Importance Sampling for rare events and conditioned random walks

M. Broniatowski[1], Y.Ritov[2]
[1]LSTA, Université Paris 6, France
michel.broniatowski@upmc.fr
[2]Dpt of Statistics, Hebrew University, Jerusalem Israël
yaacov.ritov@gmail.com

October 2009

Abstract

This paper introduces a new Importance Sampling scheme, called Adaptive Twisted Importance Sampling, which is adequate for the improved estimation of rare event probabilities in he range of moderate deviations pertaining to the empirical mean of real i.i.d. summands. It is based on a sharp approximation of the density of long runs extracted from a random walk conditioned on its end value.

Contents

0.1 Introduction and notation

Importance Sampling procedures aim at reducing the calculation time which is necessary in order to evaluate integrals, often in large dimension. We consider the case when the integral to be numerically computed is the probability of an event defined by a large number of random components; this case has received quite a lot of attention, above all when the event is of *small* probability, typically of order 10^{-8} or so, as occurs frequently in industrial applications or in communication devices. The order of magnitude of the probability to be estimated is here somehow larger, and aims at coping with "moderate probabilities" as dealt with in statistics. The basic situation in IS can be stated as follows.

Let $\mathbf{Z}$ be some random variable, say on $\mathbb{R}$, with probability measure P and density p. Let A be a subset of $\mathbb{R}$ with $P(A) > 0$. Let $Z_1^L := (Z_1, ..., Z_L)$ denote a sample of i.i.d. observations of $\mathbf{Z}$. By the law of large numbers

$$P_L := \frac{1}{L} \sum_{l=1}^{L} \mathbf{1}_A(Z_l) \tag{1}$$

estimates $P(A)$ without bias, when the $Z_i's$ are sampled under the density p. An altenative unbiased estimate of $P(A)$ can be defined through

$$P_L^g := \frac{1}{L} \sum_{l=1}^{L} \frac{p(Y_l)}{g(Y_l)} \mathbf{1}_A(Y_l) \tag{2}$$

for all density g when the support of p is a subset of the support of g, and the $Y_i's$.are i.i.d. observations of a r.v. $\mathbf{Y}$ with density g. As is well known the optimal choice for the IS sampling density g is $p_{\mathbf{Z}/A}$, the density of $\mathbf{Z}$ conditioned upon the event $(\mathbf{Z} \in A)$, unfortunately an unpracticable choice which presumes the knowledge of $P(A)$, the quantity to be estimated. Would this sampling density be at hand, the required number L of replications of $\mathbf{Y}$ to be performed would reduce to 1 and the estimate would be exactly $P(A)$. This fact motivates efforts in order to approximate $p_{\mathbf{Z}/A}$ in the case when the variable $\mathbf{Z}$ has a distribution which allows it. Sometimes the random variable $\mathbf{Z}$ is obtained as a function of a large number of random variables, say $\mathbf{X}_1^n := (\mathbf{X}_1, ..., \mathbf{X}_n)$ and the event $(\mathbf{Z} \in A)$ is of small or moderate probability. Also the density of $\mathbf{Z}$ cannot be evaluated analytically, due to the very definition of $\mathbf{Z}$, but the random variables $\mathbf{X}_i$'s have known distribution. This happens for instance when $\mathbf{Z}$ is a moment estimator or when it is the linear part of the expansion of an M or L -estimate (see Section 4). The example which we have in mind is the following, which helps as a benchmark case in the IS literature.

The r.v's $\mathbf{X}_i's$ are i.i.d. , are centered with variance 1, with common density $p_{\mathbf{X}}$ on $\mathbb{R}$, and

$$\mathbf{Z} := \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i =: \frac{1}{n} \mathbf{S}_1^n$$

is the empirical mean of the $\mathbf{X}_i's$. The set A is

$$A := (a_n, \infty) \tag{3}$$

where a_n tends slowly to $E(\mathbf{X}_1)$ from above and we intend to estimate

$$P_n := P\left(\frac{1}{n} \mathbf{S}_1^n \in A\right)$$

for large but fixed n. Many asymptotic results provide sharp estimates for $P(\mathbf{Z} \in A)$ but it is a known fact that asymptotic expansions are not always good tools when dealing with numerical approximations for fixed (even large)

n. For example, citing Ermakov (2004, p 624, [7]), the Berry-Esseen approximation for the evaluation of risks of order 10^{-2} in testing is pertinent for sample sizes of order 5000-10000; also the accuracy of available moderate deviation probabilities as developped by Inglot, Kallenberg and Ledwina in [11] has not been investigated. This motivates our interest in numerical techniques in this field.

According to (1) the basic estimate of $P(\mathbf{Z} \in A)$ is defined as follows: generate L i.i.d. samples $X_1^n(l)$ with underlying density p_X and define

$$P^{(n)}(\mathcal{E}_n) := \frac{1}{L}\sum_{l=1}^{L} \mathbf{1}_{\mathcal{E}_n}\left(X_1^n(l)\right)$$

where

$$\mathcal{E}_n := \left\{(x_1,...,x_n) \in \mathbb{R}^n : s_1^n/n > a_n\right\}. \tag{4}$$

Here $s_1^n := x_1+...+x_n$. The statistics $P^{(n)}(\mathcal{E}_n)$ estimates the *moderate deviation* probability of the sample mean of the $\mathbf{X}_i's$. Also denoting g a sampling density of the vector Y_1^n the associated IS estimate is

$$P_g^{(n)}(\mathcal{E}) := \frac{1}{L}\sum_{l=1}^{L} \frac{p_X\left(Y_1^n(l)\right)}{g\left(Y_1^n(l)\right)} \mathbf{1}_{\mathcal{E}}\left(Y_1^n(l)\right). \tag{5}$$

In the range of moderate deviations the two major contributions to IS schemes for the estimation of P_n are Fuh and Hu [10] and Ermakov [8]. The paper by Fuh and Hu does not consider events of moderate deviations as intended here; it focuses on IS schemes for the estimation of $P(Z \in A)$ where Z is a given multinormal random vector and A is a fixed set in $\mathbb{R}^d$. The authors consider efficiency with respect to the variance of the estimate and state that for the case of interest the efficient sampling scheme is deduced from the distribution of Z by a shift in the mean inside the set A. The papers by Ermakov instead handle similar problems as we do. Ermakov's 2007 paper [8] considers a sampling scheme where g is the density of i.i.d. components. He proves that this scheme is efficient in the sense that the computational burden necessary to obtain a relative precision of the estimate with respect to P_n does not grow exponentially as a function of n. He considers statistics of greater generality than the sample mean, such as M and L estimators; in the range of moderate deviations the asymptotic behavior of those objects is captured however through their linear part which is the empirical mean of their influence function, which puts the basic situation back at the center of the scene. We discuss efficiency in Section 3 and present some results in connection with Ermakov's pertaining to M and L estimators in Section 4.

The numerator in the expression (5) is the product of the $p_{\mathbf{X}_1}(Y_i)'s$ while the denominator need not be a density of i.i.d. copies evaluated on the $Y_i's$. Indeed the optimal choice for g is the density of $\mathbf{X}_1^n$ conditioned upon $\mathcal{E}_n$, say $p_{\mathbf{X}_1^n/\mathcal{E}_n}$.

Since the optimal solution is known to be $p_{\mathbf{X}_1^n/\mathcal{E}_n}$, the best its approximation, the best the sampling scheme, at least when it does not impose a large calculation burden; classical sampling schemes consist in simulation of independent copies of r.v.'s $Y_i(l)$,$1 \leq i \leq n$, and efficiency is defined in terms of variance of the estimate inside this class of sampling, which, by nature, is suboptimal with respect to sampling under good approximations of $p_{\mathbf{X}_1^k/\mathcal{E}_n}$ for long runs, i.e. for large $k = k_n$.The present paper explores the choice of good sampling schemes from this standpoint. Obviously mimicking the optimal scheme results in a net gain on the number L of replications of the runs which are necessary to obtain a given accuracy of the estimate with respect to P_n . However the criterion which we consider is different from the variance, and results as an evaluation of the MSE of our estimate on specific subsets of the runs generated by the sampling scheme, which we call typical subsets, namely having probability going to 1 under the sampling scheme as n increases. On such sets, the MSE is proved to be of very small order with respect to the variance of the classical estimate, whose MSE cannot be diminuished on any such typical subsets. We believe that this definition makes sense and prove it also numerically. This is the scope of Section 3 in which it will be shown that the relative gain in terms of simulation runs necessary to perform an $\alpha\%$ relative error on P_n drops by a factor $\sqrt{n-k}/\sqrt{n}$ with respect to the classical IS scheme.

Our proposal therefore hinges on the local approximation of the conditional distribution of longs runs $\mathbf{X}_1^k$ from $\mathbf{X}_1^n$. This cannot be achieved through the classical theory of moderate deviations, first developped by De Acosta and more recently by Ermakov; at the contrary the ad hoc procedure developped in the range of large deviations by Diaconis and Freedman [6] for the local approximation of the conditional distribution of $\mathbf{X}_1^k$ given the value of $\mathbf{S}_1^n$ is the starting point of the present approach. We find it useful to briefly expose these two different points of view. We also mention the approximation technique for moderate deviations of sub linear functionals of the empirical measure by Inglot, Kallenberg and Ledwina [11], based on strong approximation techniques; these results provide explicit equivalents for the probability of moderate deviations, but do not lead to adequate approximations for the obtention of their numerical counterparts by IS methods.

The following notation and assumptions will be kept throughout this paper.

We assume that $\mathbf{X}_1$ satisfies the Cramer condition, i.e. $\mathbf{X}_1$ has a finite moment generating function $\Phi(t) := E \exp t\mathbf{X}_1$ in a non void neighborhood of 0; denote

$$m(t) := \frac{d}{dt} \log \Phi(t) \tag{6}$$

and

$$s^2(t) := \frac{d}{dt} m(t) \tag{7}$$

when defined. The values of $m(t^\alpha) := \frac{d}{dt} \log \Phi(t^\alpha)$ and $s^2(t^\alpha) := \frac{d}{dt} m(t^\alpha)$ are the expectation and the variance of the *tilted* density

$$\pi^\alpha(x) := \frac{\exp t^\alpha x}{\Phi(t^\alpha)} p(x) \tag{8}$$

where t^α is the only solution of the equation $m(t) = \alpha$ when α belongs to the support of p. Denote Π^α the probability measure with density π^α. The *Chernoff function* of $\mathbf{X}_1$ is

$$I(x) := \sup_t tx - \log \Phi(t)$$

for x in the support of $\mathbf{X}_1$ and it holds

$$\begin{aligned} \frac{d}{dx} I(x) &= m^{\leftarrow}(x) \\ \frac{d^2}{dx^2} I(x) &= \frac{1}{s^2 \circ m^{\leftarrow}(x)} \end{aligned}$$

where $m^{\leftarrow}(x)$ denotes the reciprocal function of m.

Denote

$$\varphi(s) := \int_{-\infty}^{+\infty} e^{isx} p_{\mathbf{X}}(x) dx$$

the characteristic function of $\mathbf{X}_1$. Assume that

$$\int_{-\infty}^{+\infty} |\varphi(s)|^\nu \, ds < \infty \tag{9}$$

for some $\nu \geq 1$. This condition entails the validity of the Edgeworth expansions to be used in the sequel (see e.g. Feller [9]).

The notation $p(\mathbf{X} = x)$ is used to denote the value of the density p of the r.v. $\mathbf{X}$ at point x. The notation $p(\mathbf{S}_1^n = s)$ is used to define the value of the density of the r.v. $\mathbf{S}_1^n$ under p, i.e. when the summands are i.i.d. with density p. Also we may write $p\left(f\left(\mathbf{X}_1^n\right) = u\right)$ to denote the density (on the corresponding image space) of some function f of the sample $\mathbf{X}_1^n$. We write $\mathfrak{P}_n$ the distribution of $\mathbf{X}_1^n$ given $\mathcal{E}_n$ and $\mathfrak{p}_n$ its density. The symbol $\mathfrak{n}$ denotes the standard normal density on $\mathbb{R}$.

0.1.1 From moderate deviations to conditional distributions

A basic requirement for a good IS sampling scheme is that it mimicks the conditional *density* $p_{\mathbf{X}_1^n/\mathcal{E}_n}$. We first expose a general argument in this direction in order to clarify that there is no bypass through the general theory of large or moderate deviations to achieve this goal. Also the present discussion motivates the choices of classical IS sampling schemes (Ermakov), emphasizing that the general theory provides the proof that the *marginal* conditional distribution of $\mathbf{X}_1^n$ under $\mathcal{E}_n$ is well approximated by Π^{a_n} a statement which is usually refered to as a *Gibbs conditional principle.* We need some tools from the moderate deviation principle as developped by [7] following [5].

Let F be a class of measurable functions defined on $\mathbb{R}$ and M_F be the class of all signed finite measures on $\mathbb{R}$ which satisfy

$$\int |f| \, d\,|Q| < \infty \text{ for all } f \text{ in } F.$$

On M_F define the τ_F topology, which is the coarsest for which all mappings $f \to \int f dQ$ ($Q \in M_F$) are continuous for all f in F. For P a probability measure and Q in $M(\mathbb{R})$ the so-called Chi-square distance between P and Q is defined through

$$\chi^2(Q,P) := \frac{1}{2}\int \left(\frac{dQ}{dP}-1\right)^2 dP$$

whenever Q is absolutely continuous with respect to P, and equals $+\infty$ otherwise.

The following *moderate deviation* Sanov result holds; see [8]. Assume that a_n tends to 0 and $a_n\sqrt{n}$ tends to infinity.

Let $\mathbf{P}_n := \frac{1}{n}\sum_{i=1}^{n}\delta_{\mathbf{X}_i}$ denote the empirical measure pertaining to an i.i.d. sample $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$. Write $\mathbf{M}_n := \frac{1}{a_n}(\mathbf{P}_n - P)$. It holds

$$\begin{aligned} -\inf_{Q\in int(B)}\chi^2(Q,P) &\leq \liminf_n \frac{1}{na_n^2}\log \Pr(\mathbf{M}_n \in B) \\ &\leq \limsup_n \frac{1}{na_n^2}\log \Pr(\mathbf{M}_n \in B) \leq -\inf_{Q\in cl(B)}\chi^2(Q,P) \end{aligned} \tag{10}$$

where the interior and closure of the set B refer to the τ_F topology on M_F.

Consider now the asymptotic distribution of $\mathbf{X}_1$ conditionally upon the sequence of events $(\mathbf{S}_1^n/n > a_n x)$, so-called moderate deviation events. With $F := B(\mathbb{R}) \cup (v \to v)$ and $B(\mathbb{R})$ the class of all bounded measurable functions, (10) holds with B substitued by Ω_x the subset of M_F defined through

$$\Omega_x := \left\{ Q : \int t dQ(t) \geq x \text{ and } \int dQ(t) = 0 \right\}.$$

With P the probability measure of the r.v. $\mathbf{X}_1$ denote Q^* the χ^2 *projection of* P on Ω_x, namely

$$Q^* := \arg\inf\left\{\chi^2(Q,P), Q \in \Omega_x\right\}.$$

The set Ω_x is closed in $M_F(\mathbb{R})$ equipped with the τ_F topology. Existence of a χ^2 projection of P on a τ_F-closed subset of $M(\mathbb{R})$ holds as a consequence of Theorem 2.6 in [3] when $\int |f|\, dP$ is finite for all f in F, which clearly holds since $E\,|\mathbf{X}_1|$ is finite. Uniqueness follows from the convexity of Ω_x and the strict convexity of $Q \to \chi^2(Q,P)$. From (10) it can easily be obtained that

$$P(\mathbf{X}_1 \in A/\mathcal{E}_{n,x}) = P(A) + a_n x Q^*(A) + o(a_n) \tag{11}$$

with $\mathcal{E}_{n,x} := (\mathbf{S}_1^n \geq n a_n x)$ which in turn yields the following

Proposition 1 *With the above notation*

$$P(\mathbf{X}_1 \in A/\mathcal{E}_{n,x}) = \int_A \pi^{a_n x}(y)dy + o(1) \tag{12}$$

The proofs of (11) and of the above Proposition are differed to the appendix. This way cannot provide an equivalent expression for the conditional *density* of $\mathbf{X}_1$ which requires strong regularity assumptions. Furthermore it cannot be extended to the case of interest here, when $\mathbf{X}_1$ is substituted by $\mathbf{X}_1^k$ for large values of $k = k_n$, i.e. when an approximation of the law of the path $\mathbf{X}_1^n$ is needed, at least on long runs.

However the result in Proposition 1 is a strong argument in favor of Ermakov's sampling scheme, namely simulating i.i.d. r.v.'s with common density π^{a_n} in (5).

0.1.2 Density of a partial path conditioned on the exact value of the sum

The other way follows Zabell [16] and Diaconis and Freedman [6] approaches, which were developped in the range of large deviations. See also van Camperhout and Cover [15], who considered the density or the c.d.f. of $\mathbf{X}_1^k$ conditioned on the value of $\mathbf{S}_1^n$ for fixed k. It is restricted in essence to the context of the sample mean. The sketch of the method is as follows.

The density of $\mathbf{X}_1$ given $\mathbf{S}_1^n = ns$ writes

$$p_{\mathbf{X}_1/\mathbf{S}_1^n=ns}(x_1) = \frac{p_{\mathbf{S}_2^n}(ns-x_1)}{p_{\mathbf{S}_1^n}(ns)} p_{\mathbf{X}_1}(x_1) \tag{13}$$

where we used the symbol p to emphasize that the $\mathbf{X}_i's$ are i.i.d. with common density $p_{\mathbf{X}_1}$. It is a known fact, and easy to establish, that the density defined in (13) is invariant when sampling from any density of the form (8) instead of $p_{\mathbf{X}_1}$. This yields, selecting $\alpha = s$

$$p_{\mathbf{X}_1/\mathbf{S}_1^n=na_n}(x_1) = \frac{\pi^s_{\mathbf{S}_2^n}(ns-x_1)}{\pi^s_{\mathbf{S}_1^n}(ns)} \pi^s_{\mathbf{X}_1}(x_1).$$

When the r.v's $\mathbf{X}_i$'s obey a local central limit theorem under the sampling density $\pi^s_{X_1}$ it can be proved that

$$p_{\mathbf{X}_1/\mathbf{S}_1^n=ns}(x_1) = \pi^s_{\mathbf{X}_1}(x_1)(1+o(1)) \tag{14}$$

as n tends to ∞. Diaconis and Freedman obtain such a statement when $\mathbf{X}_1$ is substituted by $\mathbf{X}_1^k$ with $k/n \to \theta, 0 \leq \theta < 1$. We will continue this approach in the range of moderate deviations, enhancing it to the density of $\mathbf{X}_1^k$ with $k/n \to 1$. Integrating with respect to the conditional distribution of $\mathbf{S}_1^n$ under the event $\mathcal{E}_n$ provides the required approximation.

The scope of the present paper is to present some technique which provides typical realisations of runs $\mathbf{X}_1^k$ under the conditional event $\mathcal{E}_n$ for very large k. Therefore it aims at the exploration of the support of the distribution of $\mathbf{X}_1^n$

under $\mathcal{E}_n$. The application which is presented pertains to Importance Sampling for the estimation of rare events probabilities through the Adaptive Twisted IS scheme.

Section 2 of this paper is devoted to the approximation of the conditional density of $\mathbf{X}_1^k$ under $\mathcal{E}_n$. Section 3 presents the ATIS algorithm , a number of remarks for its practical implementation, and discusses efficiency . Section 4 is devoted to M and L estimates and their moderate deviation probabilities. We have postponed many proofs to the Appendix, but the main one of Section 2.

0.2 Conditioned random walks

0.2.1 Three basic Lemmas

Moderate deviations results for sums of i.i.d. real valued random variables under our assumptions have been studied since the 50's by many authors. We will make use of a *local* result, due to Richter [13], which we state as

Lemma 2 *Under the general hypotheses and notation of this paper, when* a_n *is a sequence satisfying* $\lim_{n\to\infty} a_n = 0$ *together with* $\sqrt{n}a_n \to \infty$ *it holds*

$$p\left(\frac{\mathbf{S}_1^n}{n} = a_n\right) = \frac{\sqrt{n}\exp -nI(a_n)}{\sqrt{2\pi}}\left(1 + O(a_n)\right).$$

The *global* counterpart of Lemma 2 in the form used here is due to Jensen (see [12], corollary 6.4.1) and states

Lemma 3 *Under the same hypotheses as above*

$$P\left(\frac{\mathbf{S}_1^n}{n} > a_n\right) = \frac{\exp -nI(a_n)}{\sqrt{2\pi}\sqrt{n}\psi(a_n)}\left(1 + O(\frac{1}{\sqrt{n}})\right)$$

where $\psi(a_n) := t_{a_n} s(t_{a_n})$.

The following known fact is used repetedly. It sets that the conditional densities of sub-partial sums given the partial sum is invariant through any tilting. Assume $\mathbf{X}_1, ..., \mathbf{X}_n$ i.i.d. with density p and note π^a the corresponding tilted density for some parameter a.

Lemma 4 *For* $1 \leq i \leq j \leq n$, *for all* a *in the support of* P, *for all* u *and* s

$$p\left(\mathbf{S}_i^j = u / \mathbf{S}_1^n = s\right) = \pi^a \left(\mathbf{S}_i^j = u / \mathbf{S}_1^n = s\right).$$

0.2.2 Typical paths conditioned on their sum

The sequence of constants a_n defining A in (3) and (4) satisfies

$$(A) \qquad \begin{cases} \lim_{n\rightarrow\infty} a_n\sqrt{n-k} = \infty \\ \lim_{n\rightarrow\infty} \frac{na_n}{\sqrt{n-k}} = \infty \\ \lim_{n\rightarrow\infty} a_n \left(\log n\right)^{2+\delta} = 0 \text{ for some positive } \delta \\ \lim_{n\rightarrow\infty} \frac{n-k}{n} = 0 \end{cases}$$

In this section we obtain a close approximation for $p\left(\mathbf{X}_1^k = Y_1^k/\frac{\mathbf{S}_1^n}{n} = \sigma\right)$ for $k = k_n$ where a_n and k satisfy the following set of conditions.

The value of $\frac{\mathbf{S}_1^n}{n}$ satisfies $a_n \leq \sigma \leq a_n + c_n$ with

$$(C) \qquad \begin{cases} \lim_{n\rightarrow\infty} na_nc_n = \infty \\ \lim_{n\rightarrow\infty} \frac{nc_n}{\sqrt{n-k}} = 0 \\ \lim_{n\rightarrow\infty} \frac{nc_n}{a_n(n-k)} = 0 \\ \lim_{n\rightarrow\infty} \frac{\exp -na_nc_n}{a_n(\log n)^{2+\delta}} = 0 \end{cases}$$

We denote (A1),...,(A4) , (C1),...,(C4) the above conditions.

It appears clearly from (5) that the optimal choice $g = p_{\mathbf{X}_1^n/\mathbf{S}_1^n > na_n}$ need only to hold on paths Y_1^n sampled under g and not on all $\mathbb{R}^n$. In a similar way the approximation of the optimal density need to be realized only when evaluated on samples $Y_1^n(l)$ generated according to this approximation, and approximation of $p_{\mathbf{X}_1^n/\mathcal{E}}$ on the entire space $\mathbb{R}^n$ is not needed. The approximation of $\mathfrak{p}_n$ by such a density g_n is difficult to obtain on realizations under g_n and much easier under $p_{\mathbf{X}_1^n/\mathbf{S}_1^n > na_n}$. The following Lemma proves that approximating $\mathfrak{p}_n$ by g_n under $\mathfrak{p}_n$ is similar to approximating $\mathfrak{p}_n$ by g_n under g_n.

Let $\mathfrak{R}_n$ and $\mathfrak{S}_n$ denote two p.m's on $\mathbb{R}^n$ with respective densities $\mathfrak{r}_n$ and $\mathfrak{s}_n$.

Lemma 5 *Suppose that for some sequence ε_n which tends to 0 as n tends to infinity*

$$\mathfrak{r}_n\left(Y_1^n\right) = \mathfrak{s}_n\left(Y_1^n\right)\left(1 + o_{\mathfrak{R}_n}(\varepsilon_n)\right) \tag{15}$$

as n tends to ∞. Then

$$\mathfrak{s}_n\left(Y_1^n\right) = \mathfrak{r}_n\left(Y_1^n\right)\left(1 + o_{\mathfrak{S}_n}(\varepsilon_n)\right). \tag{16}$$

Proof. Denote

$$A_{n,\varepsilon_n} := \left\{y_1^n : (1-\varepsilon_n)\mathfrak{s}_n\left(y_1^n\right) \leq \mathfrak{r}_n\left(y_1^n\right) \leq \mathfrak{s}_n\left(y_1^n\right)(1+\varepsilon_n)\right\}.$$

It holds for all positive δ

$$\lim_{n\rightarrow\infty} I(n,\delta) = 1$$

where

$$I(n,\delta) := \int \mathbf{1}_{A_{n,\delta\varepsilon_n}}\left(y_1^n\right)\frac{\mathfrak{r}_n\left(y_1^n\right)}{\mathfrak{s}_n(y_1^n)}\mathfrak{s}_n(y_1^n)dy_1^n.$$

Since

$$I(n,\delta) \leq (1+\delta\varepsilon_n)\mathfrak{S}_n\left(A_{n,\delta\varepsilon_n}\right)$$

it follows that

$$\lim_{n\to\infty} \mathfrak{S}_n\left(A_{n,\delta\varepsilon_n}\right) = 1,$$

which proves the claim. ■

This shows that the approximation of $\mathfrak{p}_n$ need not to be achieved on the whole space $\mathbb{R}^n$ but only on *typical paths* under the conditionning event $\mathcal{E}_n$. It appears that such a sharp approximation is possible on quite long portions Y_1^k of sample paths generated under $\mathfrak{P}_n$, when k tends to ∞ together with n and k/n goes to 1.

Let σ such that $a_n \leq \sigma \leq b_n$ with $b_n - a_n$ small enough. We prove that the sequence of conditional densities $p\left(\mathbf{X}_1^k = Y_1^k / \mathbf{S}_1^n = n\sigma\right)$ is closely approximated by a sequence of suitably modified tilted densities when evaluated at Y_1^k, a realization under the density $\mathfrak{p}_n$. This is the scope of Proposition 8 hereunder. The size of $b_n - a_n$ is such that $\mathfrak{p}_n\left(Y_1^k\right)$ can be substituted by an integral of $p\left(\mathbf{X}_1^k = Y_1^k / \mathbf{S}_1^n = n\sigma\right)$ with respect to the distribution of $\mathbf{S}_1^n$ conditionally on $\left(\mathbf{S}_1^n \in (na_n, nb_n)\right)$. This is the scope of Proposition 15.

Define $\Sigma_1^i := Y_1 + ... + Y_i$ and $t_{i,n}$ through

$$m(t_{i,n}) = m_{i,n} := \frac{n}{n-i}\left(\sigma - \frac{\Sigma_1^i}{n}\right) \tag{17}$$

$$s_{i,n}^2 := \frac{d^2}{dt^2}\left(\log E_{\pi^{m_{i,n}}} \exp t\mathbf{X}_1\right)(0)$$

and

$$\mu_3^{(i,n)} := \frac{d^3}{dt^3}\left(\log E_{\pi^{m_{i,n}}} \exp t\mathbf{X}_1\right)(0)$$

which are the variance and the kurtosis of $\pi^{m_{i,n}}$, reflecting the corresponding characteristics of p, since $t_{i,n}$ is close to 0 as shown in the following result.

Lemma 6 *Let σ belong to (a_n, b_n) and assume that (A) holds together with (C2) and (C3). Then under $\mathfrak{P}_n$, $t_{i,n}$ tends to 0, $s_{i,n}^2$ tends to 1 and $\mu_3^{(i,n)}$ tends to the third centered moment of p uniformly upon σ in (a_n, b_n).*

Proof. Write

$$m(t_{i,n}) = \frac{n}{n-i}(\sigma - a_n) + \frac{n}{n-i}\left(a_n - \frac{\Sigma_1^i}{n}\right)$$

which goes to 0 under $\mathfrak{P}_n$ uniformly upon σ under (C2) and (C3) where we used Lemma **??**; therefore $t_{i,n}$ goes to 0 uniformly in σ which concludes the proof. ■

The following density $g_\sigma(y_1^k)$ defined in (20) on $\mathbb{R}^k$ provides the sharp approximation of $p(\mathbf{X}_1^k = y_1^k / \mathbf{S}_1^n = n\sigma)$. This density is defined on $\mathbb{R}^k$ as a product of conditional densities which are set in the following displays. It only approximates $p(\mathbf{X}_1^k = y_1^k/\mathbf{S}_1^n = n\sigma)$ on typical vectors y_1^k which are realizations of $\mathbf{X}_1^k$ under $\mathcal{E}_n$. Chose any density $g_0(y_1)$ (for convenience denoted $g_0(y_1/y_0)$ in (20).and for $1 \leq i \leq k-1$ define recursively the sequence of conditional densities $g_i(y_{i+1}/y_1^i)$ through

$$g_0(y_1) = \pi^\sigma(y_1)$$

and

$$g_i(y_{i+1}/y_1^i) = \frac{\exp\left(y_{i+1}\left(t_{i,n} + \frac{\mu_3^{(i,n)}}{2s_{i,n}^4(n-i-1)}\right) - y_{i+1}^2/\left(2s_{i,n}^2\,(n-i-1)\right)\right)p(y_{i+1})}{K_i(y_1^i)} \tag{18}$$

a density on $\mathbb{R}$, with $t_{i,n}$ the unique solution of the equation

$$m(t_{i,n}) = \frac{n}{n-i}\left(\sigma - \frac{s_1^i}{n}\right)$$

where $s_1^i := y_1 + ... + y_i$. The normalizing factor $K_i(y_1^i)$ is

$$K_i(y_1^i) = \int \exp\left(x\left(t_{i,n} + \frac{\mu_3^{(i,n)}}{2s_{i,n}^4\,(n-i-1)}\right) - x^2/\left(2s_{i,n}^2\,(n-i-1)\right)\right)p(x)dx. \tag{19}$$

Define g_σ the density on $\mathbb{R}^k$ through

$$g_\sigma(y_1^k) := \prod_{i=0}^{k-1} g_i(y_{i+1}/y_1^i). \tag{20}$$

The definition in,(18) can also be stated as

$$g_{i+1}(y_{i+1}/x_1^i) = C_i p(y_{i+1})\mathfrak{n}\left(ab, a, y_{i+1}\right)$$

where $\mathfrak{n}\left(\mu, \sigma^2, x\right)$ is the normal density with mean μ and variance σ^2 at x. Here

$$a = s_{i,n}^2\,(n-i-1)$$

$$b = t_{i,n} + \frac{\mu_3^{(i,n)}}{2s_{i,n}^4\,(n-i-1)}$$

and the constant C_i is $\left(K_i(y_1^i)\right)^{-1}$. This form is appropriate for the simulation.

The density $g_i(y_{i+1}/y_1^i)$ is a slight modification from $\pi^{m(t_{i,n})}$. It approximates sharply $\mathrm{p}\left(\mathbf{X}_{i+1} = y_{i+1}/\mathbf{S}_1^n = n\sigma, y_1^i\right)$. For small values of i, the contribution of $y_{i+1}\frac{\mu_3^{(i,n)}}{2s_{i,n}^4(n-i-1)}$ and of $y_{i+1}^2/\left(2s_{i,n}^2\,(n-i-1)\right)$ is small and $g_i(y_{i+1}/y_1^i)$ fits nearly with $\pi^{a_n}(y_{i+1})$, when σ is close to a_n, which is in accordance both with Diaconis and Freedman's approximation when translated in the moderate deviation range and with Ermakov's IS scheme.

Remark 7 *When the X_i's are i.i.d. normal then $g_i(y_{i+1}/y_1^i) = p(y_{i+1}/y_1^i, \frac{S_n}{n} = \sigma)$ for all i.*

We then have

Proposition 8 *Set σ with $a_n \leq \sigma \leq b_n$ and assume (A) together with (C2) and (C3). Let Y_1^n be a sample with distribution $\mathfrak{P}_n$. Then uniformly upon σ*

$$p(\mathbf{X}_1^k = Y_1^k/\mathbf{S}_1^n = n\sigma) = g_\sigma(Y_1^k)(1 + o_{\mathfrak{P}_n}(a_n (\log n)^{2+\delta})). \tag{21}$$

Proof. The proof uses a Bayes formula to write $p(\mathbf{X}_1^k = Y_1^k/\mathbf{S}_1^n = n\sigma)$ as a product of k conditional densities of individual terms of the trajectory evaluated at Y_1^k, and the invariance property stated in Lemma 4. Each term of this product is approximated through an Edgeworth expansion which together with the three preceeding lemmas, conclude the proof. It holds

$$\begin{aligned} p(\mathbf{X}_1^k &= Y_1^k/\mathbf{S}_1^n = n\sigma) = p(\mathbf{X}_1 = Y_1/\mathbf{S}_1^n = n\sigma) && (22) \\ \prod_{i=1}^{k-1} p(\mathbf{X}_{i+1} &= Y_{i+1}/\mathbf{X}_1^i = Y_1^i, \mathbf{S}_1^n = n\sigma) && (23) \\ &= \prod_{i=0}^{k-1} p\left(\mathbf{X}_{i+1} = Y_{i+1}/\mathbf{S}_{i+1}^n = n\sigma - \Sigma_1^i\right) \end{aligned}$$

by the independence of the r.v's $\mathbf{X}_i$; we have set $S_1^0 := 0$. By Lemma 4

$$\begin{aligned} & p\left(\mathbf{X}_{i+1} = Y_{i+1}/\mathbf{S}_{i+1}^n = n\sigma - \Sigma_1^i\right) \\ = \; & \pi^{m_{i,n}}\left(\mathbf{X}_{i+1} = Y_{i+1}/\mathbf{S}_{i+1}^n = n\sigma - \Sigma_1^i\right) \\ = \; & \pi^{m_{i,n}}\left(\mathbf{X}_{i+1} = Y_{i+1}\right) \frac{\pi^{m_{i,n}}\left(\mathbf{S}_{i+2}^n = n\sigma - \Sigma_1^{i+1}\right)}{\pi^{m_{i,n}}\left(\mathbf{S}_{i+1}^n = n\sigma - \Sigma_1^i\right)} \end{aligned}$$

where we used Bayes formula and the independence of the $\mathbf{X}_j$'s under $\pi^{m_{i,n}}$. A precise evaluation of the dominating terms in this lattest expression is needed in order to handle the product (22).

Under the sequence of densities $\pi^{m_{i,n}}$ the i.i.d. r.v's $\mathbf{X}_{i+1}, ..., \mathbf{X}_n$ define a triangular array which satisfies a local central limit theorem, and an Edgeworth expansion. Under $\pi^{m_{i,n}}$, $\mathbf{X}_{i+1}$ has expectation $m_{i,n}$ and variance $s_{i,n}^2$. Center and normalize both the numerator and denominator in the fraction which appears in the last display. Denote $\overline{\pi_{n-i-1}^{m_{i,n}}}$ the density of the normalized partial sum $\left(\mathbf{S}_{i+2}^n - (n-i-1)m_{i,n}\right) / \left(s_{i,n}\sqrt{n-i-1}\right)$ when the summands are i.i.d. with common density $\pi^{m_{i,n}}$. Hence, evaluating both $\overline{\pi_{n-i-1}^{m_{i,n}}}$ and its normal approximation at point Y_{i+1},

$$\begin{aligned} & p\left(\mathbf{X}_{i+1} = Y_{i+1}/\mathbf{S}_{i+1}^n = n\sigma - \Sigma_1^i\right) && (24) \\ = \; & \frac{\sqrt{n-i}}{\sqrt{n-i-1}} \pi^{m_{i,n}}\left(\mathbf{X}_{i+1} = Y_{i+1}\right) \frac{\overline{\pi_{n-i-1}^{m_{i,n}}}\left(\left(m_{i,n} - Y_{i+1}\right)/s_{i,n}\sqrt{n-i-1}\right)}{\overline{\pi_{n-i}^{m_{i,n}}}(0)}. \end{aligned}$$

The sequence of densities $\overline{\pi_{n-i-1}^{m_{i,n}}}$ converges pointwise to the standard normal density under the assumptions, when $n-i$ tends to infinity, i.e. when $n-k_n$ tends to infinity, and an Edgeworth expansion to the order 5 is performed for the numerator and the denominator.

Set $Z_{i+1} := \left(m_{i,n} - Y_{i+1}\right)/s_{i,n}\sqrt{n-i-1}$. Using Lemma 24 we have

$$m_{i,n} - Y_{i+1} = a_n - Y_{i+1} + \frac{n\left(\sigma - a_n\right)}{n-i-1} + O_{\mathfrak{P}_n}\left(\frac{1}{\sqrt{n-i}}\right). \tag{25}$$

It then holds

$$\begin{aligned}\overline{\pi_{n-i-1}^{m_{i,n}}}\left(Z_{i+1}\right) \quad &= \quad \mathfrak{n}(Z_{i+1})\left[\begin{array}{c} 1+\frac{1}{\sqrt{n-i-1}}P_3(Z_{i+1})+\frac{1}{n-i-1}P_4(Z_{i+1}) \\ +\frac{1}{(n-i-1)^{3/2}}P_5(Z_{i+1}) \end{array}\right] \qquad (26)\\ &\quad +O_{\mathfrak{P}_n}\left(\frac{1}{\left(n-i-1\right)^{3/2}}\right).\end{aligned}$$

We perform an expansion in $\mathfrak{n}(Z_{i+1})$ up to the order 3, with a first order term $\mathfrak{n}\left(-Y_{i+1}/\left(s_{i,n}\sqrt{n-i-1}\right)\right)$, namely

$$\begin{aligned}\mathfrak{n}(Z_{i+1}) \quad &= \quad \mathfrak{n}\left(-Y_{i+1}/\left(s_{i,n}\sqrt{n-i-1}\right)\right) \qquad (27)\\ &\quad \left(\begin{array}{c} 1-\frac{Y_{i+1}m_{i,n}}{s_{i,n}^2(n-i-1)}+\frac{m_{i,n}^2}{2s_{i,n}^2(n-i-1)}\left(\frac{Y_{i+1}^2}{s_{i,n}^2(n-i-1)}-1\right) \\ +\frac{m_{i,n}^3}{6s_{i,n}^3(n-i-1)^{3/2}}\frac{\mathfrak{n}^{(3)}\left(\frac{Y^*}{\left(s_{i,n}\sqrt{n-i-1}\right)}\right)}{\mathfrak{n}\left(-Y_{i+1}/\left(s_{i,n}\sqrt{n-i-1}\right)\right)} \end{array}\right) \qquad (28)\end{aligned}$$

where $Y^* = \frac{1}{s_{i,n}\sqrt{n-i-1}}(-Y_{i+1}+\theta m_{i,n})$ with $|\theta|<1$. Only the first order term is relevant when handling the conditional density of the sub trajectory Y_1^k.

Write

$$m_{i,n} = \frac{n}{n-i-1}\left(a_n - \frac{\Sigma_1^i}{n}\right) + \frac{n}{n-i-1}\left(\sigma - a_n\right)$$

and use Lemmas **??** and 26 to obtain

$$\begin{aligned}\frac{\left|Y_{i+1}m_{i,n}\right|}{s_{i,n}^2\left(n-i-1\right)} \quad &= \quad \frac{O_{\mathfrak{P}_n}(\log k)}{n-i-1}\left(a_n + O_{\mathfrak{P}_n}\left(\frac{1}{\sqrt{n-i}}\right)\right)\left(1+o_{\mathfrak{P}_n}(1)\right) \qquad (29)\\ &\quad + n\left(\sigma - a_n\right)\frac{O_{\mathfrak{P}_n}(\log k)}{\left(n-i-1\right)^2}\left(1+o_{\mathfrak{P}_n}(1)\right)\end{aligned}$$

and

$$\begin{aligned}\frac{m_{i,n}^2}{s_{i,n}^2\left(n-i-1\right)} \quad &= \quad \frac{1}{n-i-1}\left(a_n + O_{\mathfrak{P}_n}\left(\frac{1}{\sqrt{n-i}}\right)\right)^2\left(1+o_{\mathfrak{P}_n}(1)\right) \qquad (30)\\ &\quad + \frac{n^2}{\left(n-i-1\right)^3}\left(\sigma - a_n\right)^2\left(1+o_{\mathfrak{P}_n}(1)\right)\\ &\quad + 2\frac{n\left(\sigma-a_n\right)}{\left(n-i-1\right)^2}\left(a_n + O_{\mathfrak{P}_n}\left(\frac{1}{\sqrt{n-i}}\right)\right)\left(1+o_{\mathfrak{P}_n}(1)\right).\end{aligned}$$

where the $1 + o_{\mathfrak{P}_n}(1)$ terms stem from the convergence of $s_{i,n}^2$ to 1 by Lemma 6. Assuming (C2) it follows that

$$\frac{|Y_{i+1} m_{i,n}|}{s_{i,n}^2 (n-i-1)} = \frac{O_{\mathfrak{P}_n}(\log k)}{n-i-1}\left(a_n + O_{\mathfrak{P}_n}\left(\frac{1}{\sqrt{n-i}}\right)\right)(1 + o_{\mathfrak{P}_n}(1))$$

and

$$\frac{m_{i,n}^2}{s_{i,n}^2 (n-i-1)} = \frac{1}{n-i-1}\left(a_n + O_{\mathfrak{P}_n}\left(\frac{1}{\sqrt{n-i}}\right)\right)^2 (1 + o_{\mathfrak{P}_n}(1))$$

which yields

$$\mathfrak{n}(Z_{i+1}) = \mathfrak{n}\left(-Y_{i+1}/\left(s_{i,n}\sqrt{n-i-1}\right)\right)\left(1 + O_{\mathfrak{P}_n}\left(\frac{a_n \log k}{n-i}\right)\right). \qquad (31)$$

The Hermite polynomials depend upon the moments of the underlying density $\pi^{m_{i,n}}$. Since $\overline{\pi_1^{m_{i,n}}}$ has expectation 0 and variance 1 the terms corresponding to P_1 and P_2 vanish. Up to the order 4 the polynomials write $P_3(x) = \frac{\mu_3^{(i)}}{6(s_{i,n})^3}(x^3 - 3x)$, $P_4(x) = \frac{\mu_3^{(i,n)}}{72(s_{i,n})^6}(x^3 - 3x) + \frac{\mu_4^{(i,n)} - 3(s_{i,n})^4}{24(s_{i,n})^4}\left(x^4 + 6x^2 - 3\right)$.

In order to obtain a development of the polynomial bracket in (26) in terms of powers of $(n-i)$ only the term in x from P_3 and the constant term from P_4 are relevant. It holds

$$\begin{aligned}\frac{P_3(Z_{i+1})}{\sqrt{n-i-1}} &= -\frac{\mu_3^{(i,n)}}{2s_{i,n}^4(n-i-1)}(a_n - Y_{i+1}) - \frac{\mu_3^{(i,n)}}{2s_{i,n}^4}\frac{n(\sigma - a_n)}{(n-i-1)^2} \\ &\quad - \frac{\mu_3^{(i,n)}(m_{i,n} - Y_{i+1})^3}{6(s_{i,n})^6 (n-i-1)^2} + O_{\mathfrak{P}_n}\left(\frac{1}{(n-i)^{3/2}}\right).\end{aligned}$$

When (C3) holds then

$$\begin{aligned}\frac{P_3(Z_{i+1})}{\sqrt{n-i-1}} &= -\frac{\mu_3^{(i,n)}}{2s_{i,n}^4(n-i-1)}(a_n - Y_{i+1}) + O_{\mathfrak{P}_n}\left(\frac{1}{(n-i)^{3/2}}\right) \\ &= \frac{\mu_3^{(i,n)}}{2s_{i,n}^4(n-i-1)}Y_{i+1} + O_{\mathfrak{P}_n}\left(\frac{1}{(n-i)^{3/2}}\right) + a_n O\left(\frac{1}{n-i}\right).\end{aligned} \qquad (32)$$

For the term of order 4 it holds

$$\frac{P_4(Z_{i+1})}{n-i-1} = \frac{1}{n-i-1}\left(\frac{1}{12s_{i,n}^3}P_3(Z_{i+1}) + \frac{\mu_4^{(i,n)} - 3s_{i,n}^4}{24s_{i,n}^4(n-i-1)}\left(Z_{i+1}^4 + 6Z_{i+1}^2 - 3\right)\right). \qquad (33)$$

When (C2) and (C3) hold it follows that

$$\frac{P_4(Z_{i+1})}{n-i-1} = -\frac{\mu_4^{(i,n)} - 3s_{i,n}^4}{8s_{i,n}^4(n-i-1)} + O_{\mathfrak{P}_n}\left(\frac{1}{(n-i-1)^{3/2}}\right).$$

The fifth term in the expansion plays no role in the asymptotics, under (A). To sum up and using (A) and Lemma 26 we get

$$\overline{\pi_{n-i-1}^{m_{i,n}}}\left(Z_{i+1}\right)=\mathfrak{n}\left(-Y_{i+1}/\left(s_{i,n}\sqrt{n-i-1}\right)\right)\left(\begin{array}{c}1+\frac{\mu_3^{(i,n)}}{2s_{i,n}^4(n-i-1)}Y_{i+1}\\ -\frac{\mu_3^{(i,n)}-s_{i,n}^4}{8s_{i,n}^4(n-i-1)}+O_{\mathfrak{P}_n}\left(\frac{a_n\log n}{n-i}\right)\end{array}\right). \tag{34}$$

Turn back to (24) and do the same Edgeworth expansion in the demominator, which writes

$$\overline{\pi_{n-i}^{m_{i,n}}}(0)=\mathfrak{n}(0)\left(1-\frac{\mu_3^{(i,n)}-s_{i,n}^4}{8s_{i,n}^4(n-i)}\right)+O_{\mathfrak{P}_n}\left(\frac{1}{(n-i)^{3/2}}\right). \tag{35}$$

Summarizing and using both (32) and (33) we obtain

$$\begin{aligned} & p\left(\mathbf{X}_{i+1}=Y_{i+1}/\mathbf{S}_{i+1}^n=n\sigma-\Sigma_i^n\right) \qquad (36)\\ = & \frac{\sqrt{n-i}}{\sqrt{n-i-1}}\exp\left(Y_{i+1}\left(t_{i,n}+\frac{\mu_3^{(i,n)}}{2s_{i,n}^2\left(n-i-1\right)}\right)-Y_{i+1}^2/\left(2s_{i,n}^2\left(n-i-1\right)\right)\right)\\ & \frac{p(Y_{i+1})}{\phi(t_{i,n})}\left(1+O_{\mathfrak{P}_n}\left(\frac{a_n\log n}{n-i}\right)\right). \end{aligned}$$

The term $\exp -Y_{i+1}^2/2s_{i,n}^2(n-i-1)$ in $g_i(Y_{i+1}/Y_i)$ comes from the ratio of the two gaussian densities $\mathfrak{n}(Z_{i+1})$ and $\mathfrak{n}(0)$. Taking logarithms and using standard calculus provides the result in (18); indeed the constant term $-\frac{\mu_3^{(i,n)}-3s_{i,n}^4}{8s_{i,n}^4(n-i-1)}$ in (33) combines with the corresponding one in (35) to produce a term of order $O_{\mathfrak{P}_n}\left(\frac{1}{(n-i)^2}\right)$ whose sum is $O_{\mathfrak{P}_n}\left(\frac{1}{(n-k)}\right)=o_{\mathfrak{P}_n}\left(a_n\left(\log n\right)^{2+\delta}\right)$.

We now prove that K_i as defined in (19) satisfies

$$K_i(Y_1^i)=\phi(t_{i,n})\left(1-\frac{1}{2(n-i-1)}\right)+O_{\mathfrak{P}_n}\left(\frac{1}{(n-i)^{3/2}}\right). \tag{37}$$

This will conclude the proof.

Use the classical bounds

$$1-u+\frac{u^2}{2}-\frac{u^3}{6}\leq e^{-u}\leq 1-u+\frac{u^2}{2}$$

to obtain on both sides of the above inequalities the second order approximation of $K_i(Y_1^i)$. The upper bound is

$$\begin{aligned} K_i(Y_1^i) \quad \leq \quad & \phi\left(t_{i,n}\right)+\frac{\mu_3^{(i,n)}}{2s_{i,n}^2\left(n-i-1\right)}\phi'\left(t_{i,n}\right)+\frac{\mu_3^{(i,n)2}}{(2)^2s_{i,n}^4\left(n-i-1\right)^2}\phi\text{"}\left(t_{i,n}\right)\\ & -\frac{1}{2s_{i,n}^2(n-i-1)}\left[\phi\text{"}\left(t_{i,n}\right)+\frac{\mu_3^{(i,n)}}{2s_{i,n}^2\left(n-i-1\right)}\phi^{(3)}(t_{i,n})\right]. \end{aligned}$$

The lower bound is the same up to order 2 and the third order term plays no role.

Use Lemma 6 to conclude, making a Taylor expansion in $\phi\left(t_{i,n}\right)$, $\phi'\left(t_{i,n}\right)$ and $\phi"\left(t_{i,n}\right)$. The dominating terms are due to $\phi\left(t_{i,n}\right)$ and $\frac{1}{2s_{i,n}^2(n-i-1)}\phi"\left(t_{i,n}\right)$ which yield the $1-\frac{1}{2(n-i-1)}$ term in (37). The other terms are indeed

$$O_{\mathfrak{P}_n}\left(\frac{1}{(n-i)^{3/2}}\right)$$

using Lemma 24, leading to (37). Hence (**??**) writes as

$$(36)=g_i(Y_{i+1}/Y_1^i)\left(1+O_{\mathfrak{P}_n}\left(\frac{a_n\log n}{n-i}\right)\right).$$

Putting the pieces together yields under (A)

$$p(\mathbf{X}_1^k=Y_1^k/\mathbf{S}_1^n=n\sigma)=\left(1+o_{\mathfrak{P}_n}\left(a_n\left(\log n\right)^{2+\delta}\right)\right)\prod_{i=1}^{k}g_i(Y_{i+1}/Y_1^i).$$

Uniformity upon σ is a consequence of Lemma 6. This closes the proof of the Proposition. ■

Remark 9 *When the X_i's are i.i.d. normal, then the result in the above Proposition holds with $k=n$ stating that $p(\mathbf{X}_1^n=x_1^n/\mathbf{S}_1^n=n\sigma)=g_\sigma\left(x_1^n\right)$ for all x_1^n in $\mathbb{R}^n$.*

Remark 10 *The density in (18) is a slight modification of $\pi^{m_{i,n}}$. However second order terms are required here in order to handle the approximation of the density of $\mathbf{X}_{i+1}$ conditioned upon $\mathbf{X}_1^i$ and $\mathbf{S}_1^n/n$. The modification from $\pi^{m_{i,n}}$ to g_i is a small shift in the location parameter, which reflects the asymmetry of the underlying distribution p, and a change in the variance : large values of $\mathbf{X}_{i+1}$ have smaller weight for large i, which is to say that the distribution of $\mathbf{X}_{i+1}$ tends to concentrate around $m_{i,n}$ as i approaches k.*

Remark 11 *The "moderate deviation" case is typically $a_n=n^{-\tau}$, for τ in $(0,1/2)$. In this case the condition $a_n\left(\log n\right)^{2+\delta}\rightarrow 0$ holds for all values of τ. The other case is when a_n is "nearly constant", in the range $a_n=\left(\log n\right)^{-\gamma}$,$\gamma<2$, decreasing very slowly to 0, with $\gamma>2+\delta,\delta>0$.*

Remark 12 *In Lemmas* **??** *and 27 , as in the previous Proposition, we use an Edgeworth expansion for the density of the normalized sum of the $n-$th row of some triangular array of row-wise independent r.v's with common density. Consider the i.i.d. r.v's $\mathbf{X}_1,...,\mathbf{X}_n$ with common density $\pi^\sigma(x)$ where σ may depend on n but remains bounded. The Edgeworth expansion pertaining to $\overline{\pi_n^\sigma}$ can be derived following closely the proof given for example in [9], pp 532 and followings substituting the cumulants of p by those of π^σ. Denote $\varphi_\sigma(z)$ the characteristic function of $\pi^\sigma(x)$. Clearly for any $\delta>0$ there exists $q_{\sigma,\delta}<1$*

such that $|\varphi_\sigma(z)| < q_{\sigma,\delta}$ and since a_n is bounded, $\sup_n q_{\sigma,\delta} < 1$. Therefore the inequality (2.5) in [9] p533 holds. With ψ_n defined as in [9] (2.6) holds with φ replaced by φ_σ and σ by $s(t_\sigma)$; (2.9) holds, which completes the proof of the Edgeworth expansion in the simple case. The proof goes in the same way for higher order expansions. This justifies our argument in the Lemmas cited above. In the proofs of Proposition 8 we made use of such expansions when the r.v's $\mathbf{X}_{i+1}, ..., \mathbf{X}_n$ are i.i.d. with common density $\pi^{m_{i,n}}(x)$.The same argument as sketched hereabove applies in this case also.

0.2.3 Conditioning on final events $\mathcal{E}_n$

Let $\mathbf{T} := \mathbf{S}_1^n/n$ with distribution under the conditioning event $\mathcal{E}_n$. Hence for any Borel set A

$$P\left(\mathbf{T}\in A\right)=\mathfrak{P}_n\left(\frac{\mathbf{S}_1^n}{n}\in A\right). \tag{38}$$

The distribution of $\mathbf{T}$ is concentrated on a small neighborhood of a_n. Indeed we have

Lemma 13 *Assume that (A1) holds. For any sequence c_n such that (C1) holds,*

$$P\left(a_n \leq \mathbf{T} \leq a_n + c_n\right) = 1 + O\left(\exp -n a_n c_n\right).$$

Proof. *Use Lemma 3.* ■

Moreover $\mathbf{T}$ is asymptotically exponentially distributed. The asymptotic distribution of $\mathbf{T}$ is captured in the following

Lemma 14 *When (A1) holds then for all u in $\mathbb{R}^+$ the r.v. $\mathbf{Z} := nt^{a_n}\left(\mathbf{T} - a_n\right)$ satisfies*

$$p_{\mathbf{Z}}\left(u\right) = e^{-u}\left(1+o(1)\right)$$

where $m(t^{a_n}) = a_n$ and therefore $\mathbf{T} = a_n + O_P\left(\frac{1}{na_n}\right)$.

Proof. Write

$$p_{\mathbf{Z}}\left(u\right) = \frac{1}{nt^{a_n}} \frac{p_{\mathbf{S}_n/n}\left(a_n + u/\left(nt^{a_n}\right)\right)}{P\left(\mathbf{S}_n/n > a_n\right)}$$

and use Lemmas 2and 3. A first order expansion yields $a_n = m(t^{a_n}) = t^{a_n}\left(1+o(1)\right)$ which proves the claim. ■

In this Section Proposition 8 is extended in order to provide an approximation of $\mathfrak{p}_n(Y_1^k)$ when Y_1^k is a random vector generated under $\mathfrak{p}_n$. This is obtained through an integration w.r.t. σ in (21); indeed it holds

$$\mathfrak{p}_n(Y_1^k) := \int_{a_n}^{\infty} p\left(\mathbf{X}_1^k = Y_1^k / \mathbf{T} = \sigma\right) p_{\mathbf{T}}\left(\sigma\right) d\sigma \tag{39}$$

and the domain of integration can be reduced to a small neighborhood of a_n which contains nearly all the realizations of $\mathbf{T}$ under $\mathcal{E}_n$. This argument allows the interchange of asymptotic equivalents and integration.

Define

$$g_n(x_1^k) := \int_{a_n}^{\infty} g_\sigma(x_1^k) p_{\mathbf{T}}(\sigma)\, d\sigma$$

where $g_\sigma(x)$ is defined in (20).

When g_σ is substituted by g_n then

$$\mathfrak{p}_n(Y_1^k) = g_n(Y_1^k)\,(1 + o_{\mathfrak{P}_n}(1))$$

does not stand.

Let

$$b_n = a_n + c_n$$

where c_n is fitted compatibly with Proposition 8.

Define

$$\overline{g_n}(y_1^k) := \frac{\int_{a_n}^{b_n} g_\sigma(y_1^k) p_{\mathbf{T}}(\sigma)\, d\sigma}{P\left(\mathbf{T} \in (a_n, b_n)\right)}. \tag{40}$$

Proposition 15 *When Y_1^k is a random vector generated with density $\mathfrak{p}_n$ and (A) and (C) hold then*

$$\mathfrak{p}_n(Y_1^k) = \overline{g_n}(Y_1^k)\left(1 + o_{\mathfrak{P}_n}(a_n (\log n)^{2+\delta})\right). \tag{41}$$

The proof of Proposition 15 relies upon the following Lemma, whose proof is postponed to the Appendix.

Lemma 16 *Let b_n satisfy $b_n = a_n + c_n$ and (A) and (C) hold then when Y_1^n is generated under $\mathfrak{p}_n$ it holds*

$$\mathfrak{p}_n\left(Y_1^k\right) = \int_{a_n}^{b_n} p\left(Y_1^k / \mathbf{T} = \sigma\right) p\left(\mathbf{T} = \sigma\right) d\sigma \left(1 + O_{\mathfrak{P}_n}\left(\exp -n a_n c_n\right)\right).$$

We now prove Proposition 15 through an integration of the local approximation given in Proposition 8.

For all σ in (a_n, b_n)

$$g_\sigma\left(Y_1^k\right) = p\left(Y_1^k / \mathbf{T} = \sigma\right)\left(1 + o_{\mathfrak{P}_n}(a_n (\log n)^{2+\delta})\right) \tag{42}$$

uniformly on σ when Y_1^n is sampled under $\mathfrak{p}_n$. It then holds

$$\begin{aligned}
\overline{g_n}(Y_1^k) \quad &: \quad = \frac{\int_{a_n}^{b_n} g_\sigma\left(Y_1^k\right) p(\mathbf{T}{=}\sigma) d\sigma}{P\left(a_n < \mathbf{T} < b_n\right)} \\
&= \int_{a_n}^{b_n} p\left(Y_1^k / \mathbf{T} = \sigma\right) p(\mathbf{T} = \sigma) d\sigma \left(1 + o_{\mathfrak{P}_n}(a_n (\log n)^{2+\delta})\right) \\
&= \mathfrak{p}_n\left(Y_1^k\right)\left(1 + o_{\mathfrak{P}_n}(a_n (\log n)^{2+\delta})\right)
\end{aligned}$$

where we used Lemmas 16 together with (C4) which helps to keep the $o_{\mathfrak{P}_n}\left(a_n\left(\log n\right)^{2+\delta}\right)$ term. This concludes the proof of Proposition 15.

As a consequence of Lemma 5 the following result holds, which asseses that when sampled under $\overline{g_n}$ the likelihood of the random vector X_1^k approximates $\mathfrak{p}_n(X_1^k)$.

Proposition 17 *Assume (A) and (C). Let X_1^k be a random vector with p.m. $\overline{G_n}$ with density $\overline{g_n}$ on $\mathbb{R}^k$ defined in (40). It holds*

$$\overline{g_n}\left(X_1^k\right)=\mathfrak{p}_n(X_1^k)\left(1+o_{\overline{G_n}}(a_n\left(\log n\right)^{2+\delta})\right)$$

as $n\rightarrow\infty$.

0.3 The Adaptive Twisted Importance Sampling scheme

The last result in Proposition 17 above suggests that an Importance Sampling density deduced from $\overline{g_n}$ would benefit from some optimality as defined in the Introduction since it fits with the conditional density on long runs. It is enough to approximate the conditional distribution of $\mathbf{T}=\mathbf{S}_1^n/n$ under $\mathcal{E}_n$ by Lemma 14 and to plug in this approximation in (40).

Let $\mathbf{E}$ denote a r.v. with exponential distribution with parameter na_n on $(a_n,+\infty)$

$$p_{\mathbf{E}}(t):=na_ne^{-na_n(t-a_n)}\mathbf{1}_{(a_n,+\infty)}(t). \tag{43}$$

Using again Lemmas 2 and 3 it is easily checked that

$$\sup_{na_n\leq s\leq nb_n}\frac{p(\mathbf{E}=s)}{p(\mathbf{T}=s)}=1+o(\varepsilon_n)$$

for some sequence ε_n whinch tends to 0, from which

$$\mathbf{g}\left(X_1^k\right):=\frac{\int_{a_n}^{b_n}g_\sigma\left(X_1^k\right)p(\mathbf{E}=\sigma)d\sigma}{\int_{a_n}^{b_n}p(\mathbf{E}=\sigma)d\sigma}=\mathfrak{p}_n\left(X_1^k\right)\left(1+o_{\mathfrak{P}_n}(\varepsilon'_n)\right) \tag{44}$$

with $\lim_{n\rightarrow\infty}\varepsilon'_n=0$, which proves that we may substitute $\mathbf{T}$ by the exponential r.v. $\mathbf{E}$ while keeping the properties of the IS procedure. We denote

$$\mathbf{g}\left(x_1^n\right):=\mathbf{g}\left(x_1^k\right)\prod_{i=k+1}^{n}\pi^{\alpha_k}(x_i) \tag{45}$$

the sampling scheme under which the estimate (5) is computed; in (45) the value of α_k is defined through

$$\alpha_k:=m(t_k)$$

with

$$m(t_k) = \frac{n}{n-k}\left(a_n - \frac{s_1^k}{n}\right).$$

0.3.1 The Adaptive Twisted IS algorithm

Since the r.v. $\mathbf{E}$ is highly concentrated in a small neighborhood of a_n we suggest to forget about b_n in the definition (45) of $\mathbf{g}$ above and to integrate on (a_n, ∞) instead of (a_n, b_n). Numerical experiments argue in favor of this heuristic. The remarks at the end of this paragraph provide simple and efficient solutions for the effective calculation of the estimate.

1- Draw M independent random variables $E^1..., E^M$ with distribution (43) and define the density on $\mathbb{R}^n$

$$\overline{\mathbf{g}}(x_1^n) := \frac{1}{M}\sum_{m=1}^{M}\left(g_{E^m}(x_1^k)\prod_{i=k+1}^{n}\pi^{\alpha_k}(x_i^i)\right) \tag{46}$$

where g_{E^m} is defined as

$$g_{E^m}(x_1^k) := \prod_{i=1}^{k-1} g_{i+1}(x_{i+1}/x_1^i) \tag{47}$$

where $g_{i+1}(x_{i+1}/x_1^i)$ is defined in (18) for $i \geq 1$, $g_0(x_1) = \pi^{a_n}(x_1)$ and

$$\pi^{\alpha_k}(x) := \frac{\exp t_k x}{\Phi(t_k)}p(x) \tag{48}$$

where $\alpha_k = m(t_k)$ and t_k is the only solution of the equation

$$m(t_k) = \frac{n}{n-k}\left(a_n - \frac{s_1^k}{n}\right) \tag{49}$$

with $s_1^k := x_1 + ... + x_k$, with $s_1^0 = 0$.

2-Define L which is the number of replications of the simulated random trajectory to be performed

```
3-For $l$ between 1 and $L$ do
        {
        draw a random variable $E(l)$ with distribution (43)
        draw the first $k$ variables $X_1^k(l)$ recursively with density $g_{E(l)}(x_1^k)$ as defined in (47) with $E^m$ substituted by $E(l)$.
        Draw the $n-k$ random variables $X_{k+1}^n(l)$ independently with common density $\pi^{\alpha_k}(x)$ defined in (48) with $E^m$ substituted by $E(l)$.
}
```

4- Define

$$\widehat{P_n} := \frac{1}{L}\sum_{l=1}^{L}\frac{\prod_{i=0}^{n} p(X_i(l))}{\overline{g}(X_1^n(l))}\mathbf{1}_{\mathcal{E}_n}(l) \tag{50}$$

where

$$\mathbf{1}_{\mathcal{E}_n}(l) := \mathbf{1}_{(a_n,\infty)}\left(S_1^n(l)/n\right) \tag{51}$$

Some remarks for the implementation of the algorithm

A number of remarks hereunder show that ATIS is not difficult to implement. Since the first order efficiency of i.i.d sample schemes is reached if and only if the sampling distribution is the twisted one with parameter a_n (see [8]), the present algorithm should be compared with it. The classical IS scheme which uses i.i.d. replicates with density π^{a_n} is easy to implement but may lead to biased estimates of P_n; the simulation of a r.v. with density π^{a_n} is difficult in non standard cases When p is easy to simulate then an Acceptance/Rejection algorithm can be used; however this requires to truncate the support of p, what should precisely be avoided in order to obtain unbiaised estimates; see [2]. When π^{a_n} is easy to simulate, ATIS may take more time to run, due to the various intermediate calculations which are required at each stage of the algorithm.

The generation of the r.v. $X_1^k(l)$ above is easy and fast and does not require any simulation according to a twisted density. It holds

$$g_{i+1}(x_{i+1}/x_1^i) = C_i p(x_{i+1})\mathfrak{n}\left(ab, a, x_{i+1}\right) \tag{52}$$

where $\mathfrak{n}\left(\mu, \sigma^2, x\right)$ is the normal density with mean μ and variance σ^2 at x. Here

$$a = s_{i,n}^2\left(n-i-1\right)$$

$$b = t_{i,n} + \frac{\mu_3^{(i,n)}}{2s_{i,n}^4\left(n-i-1\right)}.$$

A r.v. Y with density $g(x) = Cp(x)n(x)$, with $C = \left(\int p(x)n(x)dx\right)^{-1}$ and where p is a given density and $n(x) = \mathfrak{n}\left(\mu, \sigma^2, x\right)$ is easy to simulate: Denote $\mathfrak{N}$ the c.d.f. with density $\mathfrak{n}\left(\mu, \sigma^2, x\right)$. It is easily checked that $g(x)$ is the density of the r.v. $Y := \mathfrak{N}^{\leftarrow}(X)$ where X is a r.v. on $[0,1]$ with density $h(u) := \frac{1}{C}p\left(\mathfrak{N}^{\leftarrow}(u)\right); \mathfrak{N}^{\leftarrow}$ denotes the reciprocal function of $\mathfrak{N}$. Now an acceptance/rejection algorithm provides a realisation of X. Indeed let $f(x)$ be a density such that $p\left(\mathfrak{N}^{\leftarrow}(u)\right) \leq Kf(x)$ for some constant K and all x in $[0,1]$; Let $\mathcal{P}$ be uniformly distributed on the hypograph of Kf, namely $\mathcal{P} := \left(X_{\mathcal{P}}, Y_{\mathcal{P}} = KUf\left(X_{\mathcal{P}}\right)\right)$ where $X_{\mathcal{P}}$ has density f and U is uniform $[0,1]$ independent of $X_{\mathcal{P}}$. When $Y_{\mathcal{P}}$ is less than $p\left(\mathfrak{N}^{\leftarrow}(X_{\mathcal{P}})\right)$ then $X_{\mathcal{P}}$ has density h.

The calculation of $\overline{\boldsymbol{g}}(X_1^n(l))$ above requires the value of $C_i = \left(\int p(x)\mathfrak{n}\left(ab, a, x\right)dx\right)^{-1}$ in (52). A Monte Carlo technique can be used: simulate N i.i.d. r.v's Z_j with density $\mathfrak{n}\left(ab, a, .\right)$, which is fast, and substitute C_i by $\widehat{C_i} := \left(\frac{1}{N}\sum_{j=1}^N p(Z_j)\right)^{-1}$, which provides a very accurate approximation to be inserted in the calculation of the estimate.

It may seem that this algorithm requires to solve Lk equations of the form $m(t) = \frac{n}{n-i}\left(\sigma - \frac{S_1^i}{n}\right)$ in order to obtain the $t_{i,n}$ which are necessary to perform the simulation of $X_1^k(l)$ as described above as well as the calculation of $\overline{\mathbf{g}}(x_1^n)$. Such is is not the case, and only L equations have to be solved. Consider for example the simulation of $X_1^k(l)$ with density $g_{E(l)}(x_1^k)$. This is achieved as follows:

1- Solve the equation

$$m(t) = E(l)$$

whose solution is $t_{0,n}$. Generate $X_0(l)$ according to $\pi^{E(l)}$.

2- Since

$$m(t_{i+1,n}) - m(t_{i,n}) = -\frac{1}{n-i}\left(m(t_{i,n}) + X_i(l)\right)$$

use a first order approximation to derive

$$t_{i+1,n} \simeq t_{i,n} - \frac{1}{(n-i)\, s(t_{i,n})}\left(m(t_{i,n}) + X_i(l)\right)$$

from which (52) is derived and $X_{i+1}(l)$ can be simulated as mentioned above. In the moderate deviation scale the function $s^2(.)$ does not vary from 1 and the above approximation is fair.

Remark 18 *The density $\overline{\mathbf{g}}(x)$ on $\mathbb{R}^n$ is a Monte Carlo approximation of g_n defined by*

$$g_n(x) := \int g_\sigma(x) p\left(\mathbf{T} = \sigma\right) d\sigma$$

where $p\left(\mathbf{T} = \sigma\right)$ is replaced by $p(\mathbf{E} = \sigma)$ and the integral is replaced by a finite mixture. M is a free parameter. Also notice that the $n-k$ i.i.d. r.v's have common tilted density $\pi^{\alpha_k}(x)$ with parameter given by (49), thus identical to Ermakov's sampling scheme with end point in$\left(a_n - \frac{s_1^k}{n}, \infty\right)$, and not in $\left(m(t_{k-1}) - \frac{s_1^k}{n}, \infty\right)$.

0.3.2 The choice of the tuning parameters

Choosing k

The critical parameter k is the length of the partial sum run which is to be simulated according to the density $\mathbf{g}(x_1^k)$ as defined in (44). By (44) it would be enough to establish some statistics averaging the estimate ratios $\mathbf{g}\left(X_1^j\right)/\mathfrak{p}_n\left(X_1^j\right)$ on a set of runs , and to select k as some j ensuring that this ratio keeps close to 1. In the case when the r.v's X_i are normally distributed the density $g_i(y_{i+1}/y_1^i)$ as defined in (18) coincides with $p(y_{i+1}/y_i, \frac{S_n}{n} = \sigma)$ for all value of i between 1 and $n-1$ which entails that k can be set equal to $n-1$. This very peculiar case is illustrated in Figure 1, for $n = 100$, and P_n is close to 0.01. We can see that ATIS produces a very sharp estimate of P_n for a small value of L when compared to the classical IS scheme.

In the other cases, when $g_i(y_{i+1}/y_1^i)$ approximates $p(y_{i+1}/y_i, \frac{S_n}{n} = \sigma)$ only under some conditions on k as described in Conditions (A), we propose the following heuristics, which works well and is easy to implement; other choices are possible, which provide similar acceptable results. Instead of $\mathbf{g}$ consider the following construction, which will also be used in the IS algorithm: simulate $E^1..., E^M$, i.i.d. with distribution (43) and define the density on $\mathbb{R}^{j+1}$

$$\overline{\mathbf{g}}(x_0^j) := \frac{1}{M}\sum_{m=1}^{M} g_{E^m}(x_1^j)\pi^{E^m}(x_0)$$

where g_{E^m} is defined as

$$g_{E^m}(x_0^j) := \prod_{i=1}^{j-1} g_{i+1}(x_{i+1}/x_1^i)$$

where $g_{i+1}(x_{i+1}/x_1^i)$ is defined in (18) for $i \geq 1$. The density $\overline{\mathbf{g}}(x_0^j)$ is a Monte Carlo approximation of $\mathbf{g}\left(x_0^j\right)$.

By (39) and following the same heuristics as for $\overline{\mathbf{g}}$ define , with a new set of i.i.d. E^m's

$$\overline{\mathfrak{p}_n}\left(x_0^j\right) := \frac{1}{M}\sum_{m=1}^{M} p\left(\mathbf{X}_0^j = x_0^j / \mathbf{T} = E^m\right).$$

We use Lemma 2 in order to obtain an explicit approximation for $\overline{\mathfrak{p}_n}$. It holds

$$\begin{aligned} p\left(\mathbf{X}_0^j = x_0^j/\mathbf{T} = E^m\right) &= \frac{p\left(\mathbf{S}_{j+1}^n = n\left(E^m - \frac{s_0^j}{n}\right)\right)}{p\left(\mathbf{S}_1^n = nE^m\right)} p\left(\mathbf{X}_0^j = x_0^j\right) \\ &= \sqrt{\frac{n-j}{n}} \frac{\exp -(n-j) I\left(\frac{n}{n-j}\left(E^m - \frac{s_0^j}{n}\right)\right)}{\exp -nI\left(E^m\right)} p\left(\mathbf{X}_0^j = x_0^j\right)(1 + o(1)). \end{aligned}$$

Define therefore

$$\left(\widehat{\mathfrak{p}_n}\right)_m\left(x_0^j\right) := \sqrt{\frac{n-j}{n}} \frac{\exp -(n-j) I\left(\frac{n}{n-j}\left(E^m - \frac{s_0^j}{n}\right)\right)}{\exp -nI\left(E^m\right)} p\left(\mathbf{X}_0^j = x_0^j\right)$$

and

$$\widehat{\mathfrak{p}_n}\left(x_0^j\right) := \frac{1}{M}\sum_{m=1}^{M}\left(\widehat{\mathfrak{p}_n}\right)_m\left(x_0^j\right).$$

Fix some integer L which is the number of runs to be simulated in order to fix k; L need not be large. For all l between 0 and L draw independently a random variable E^l with density (43) and the run $X_0^j(l)$ with density g_{E^l} defined as in (20) with k substituted by j and σ by E^l.

$$\frac{1}{L}\sum_{l=1}^{L} \frac{\overline{\mathbf{g}}(X_0^j(l))}{\widehat{\mathfrak{p}_n}\left(X_0^j(l)\right)}.$$

Fix k as the smallest j which indicates a departure of this statistics from 1.

The choice of M

In ATIS the distribution in (44) is substituted by a numerical approximation of

$$\int_{a_n}^{\infty} g_\sigma \left(X_1^k\right) p(\mathbf{E}=\sigma)d\sigma \tag{53}$$

which is suboptimal with respect to (44) but is easily implemented. A Monte Carlo procedure produces $\overline{\mathbf{g}}(x_1^n)$ as described above in (46). It appears that M should be large when k is large. For example in the normal case with $n=100$, for $k=60$, then $M=30$ produces excellent estimates for values of L of order 5000, whereas for $k=98$, the value of M should increase up to 2000, with the same L .as seen in Figure2. The reason for this increase in M is that (53) is a mixture of densities in very high dimension, which seems very sensitive with respect to the approximation of the mixture measure. This point should deserve a specific study, out of the scope of the present paper. However the normal case is quite specific, since it allows k to be as close to n as wanted. In the other cases, as examplified in the figures pertaining to the exponential case, k is resticted to lower values and M is rather low.

0.3.3 Asymptotic efficiency of the adaptive twisted IS scheme

The evaluation of the performances of IS algorithm is a controversal argument. Many criterions are at hand, for example the *probability of hits* which counts the relative number of simulations hitting the target (a_n, ∞) , or the *variance* of the estimator. We refer to the book by Bucklew [4] for a discussion on the relative merits of each approach.

The variance of an IS estimate of P_n under the sampling density g writes

$$VarP_g^{(n)}(\mathcal{E}) = \frac{1}{L}\left(E_g\left(P_g(l)\right)^2 - P_n^2\right)$$

with

$$P_g(l) := \frac{p\left(Y_1^n\left(l\right)\right)}{g\left(Y_1^n\left(l\right)\right)}\mathbf{1}_{\mathcal{E}_n}\left(\Sigma_1^n\left(l\right)\right).$$

The situation which we face with our proposal lacks the possibility to provide an order of magnitude of the variance our our IS estimate, since the properties necessary to define it have been obtained only on *typical paths* under the sampling density $\mathbf{g}$ defined in (45) and not on the whole space $\mathbb{R}^n$ (but in the case when the X_i's are normally distributed). We will prove , however, that the performance of this new procedure can be considered favorably. Not surprisingly the loss of performance with respect to the optimal sampling density $p_{\mathbf{X}_1^n/\mathcal{E}_n}$ is due to the $n-k$ last i.i.d. simulations, leading a quasi- MSE of the estimate proportional to $\sqrt{n-k}$.

In order to discuss this we first go back to the classical IS scheme, for which we evaluate the asymptotic variance.

The variance of the classical IS scheme and a discussion on efficiency

The asymptotic variance of the estimate of $P(\mathcal{E}_n)$ can be evaluated as follows.

The classical IS is defined simulating L times a random sample of n i.i.d. r.v's $X_1^n(j)$, $1 \leq j \leq L$, with tilted density π^{a_n}. The standard IS estimate is defined through

$$\overline{P_n} := \frac{1}{L}\sum_{l=1}^{L} \mathbf{1}_{\mathcal{E}_n}(l) \frac{\prod_{i=1}^n p(X_i(l))}{\prod_{i=1}^n \pi^{a_n}(X_i(l))}$$

where the $X_i(l)$ are i.i.d. with density π^{a_n} and $\mathbf{1}_{\mathcal{E}_n}(l)$ is as in (51). Set

$$\overline{P_n}(l) := \mathbf{1}_{\mathcal{E}_n}(l) \frac{\prod_{i=1}^n p(X_i(l))}{\prod_{i=1}^n \pi^{a_n}(X_i(l))}.$$

The variance of $\overline{P_n}$ is given by

$$Var\overline{P_n} = \frac{1}{L}\left(E_{\pi^{a_n}} \left(\overline{P_n}(l)\right)^2 - P_n^2 \right).$$

The *relative accuracy* of the estimate P_n^{IS} is defined through

$$RE(\overline{P_n}) := \frac{Var\overline{P_n}}{P_n^2} = \frac{1}{L}\left(\frac{E_{\pi^{a_n}} \left(\overline{P_n}(l)\right)^2}{P_n^2} - 1 \right).$$

It holds

Proposition 19 *The relative accuracy of the estimate P_n^{IS} is given by*

$$RE(\overline{P_n}) = \frac{\sqrt{2\pi}\sqrt{n}}{L} a_n(1 + o(1)) \text{ as } n \text{ tends to infinity.}$$

Proof. It holds, omitting the index l for brevity and noting a for a_n

$$\begin{aligned} E_{\pi^a} \left(\overline{P_n}(l)\right)^2 &= E_p \left(\mathbf{1}_{\mathcal{E}_n}(X_1^n) \frac{p(X_1^n)}{\pi^a(X_1^n)} \right) \\ &= \phi^n(t^a) \exp -nat^a \int_{na}^{\infty} \exp -t^a (s - na) \, p_{S_n}(s) ds. \end{aligned}$$

The Laplace integral above satisfies

$$\int_{na}^{\infty} \exp -t^a (s - na) \, p_{S_n}(s) ds = P_n(1 + o(1))$$

as n tends to infinity, which, together with the expansion

$$\phi^n(t_a) \exp -nat^a = P_n \sqrt{n}\sqrt{2\pi} t^a (1 + o(1))$$

(which holds when $\lim_{n\to\infty} a\sqrt{n} = \infty$) concludes the proof. We have used Lemma 6 to assess that $\lim_{n\to\infty} s(t^a) = 1$. ■

We now come to a discussion of the above result. It is well known that the variance is not a satisfactory criterion to describe the variability of the outcomes of a random phenomenon: for example, a sequence of symmetric r.v's $\mathbf{X}_n$ taking values $-\exp\exp n, 0, \exp\exp n$ with relative frequencies defined through $P(\mathbf{X}_n = \exp n) = \exp -n$ has variance going to ∞ while being concentrated at 0. In this case we can define an increasing family of sets B_n with $P(\mathbf{X}_n \in B_n) \to 1$ on which $E\left(\mathbf{1}_{B_n}\mathbf{X}_n^2\right) = 0$, a much better indicator, obtained through trimming. We will prove that such an indicator cannot be defined for the classical IS scheme, stating therefore that the variance rate obtained in Proposition 19 is indeed meaningfull.

The easy case when $\mathbf{X}_1, ..., \mathbf{X}_n$ are i.i.d. with standard normal distribution is sufficient for our need.

The variance of the IS estimate is proportional to

$$\begin{aligned} V \quad : \quad &= E_p \mathbf{1}_{(na_n,\infty)}\left(\mathbf{S}_1^n\right)\frac{p\left(\mathbf{X}_1^n\right)}{\pi^{a_n}\left(\mathbf{X}_1^n\right)} - P_n^2 \\ = \quad & E_p \mathbf{1}_{(na_n,\infty)}\left(\mathbf{S}_1^n\right)\left(\exp\frac{na_n^2}{2}\right)\left(\exp -a_n\mathbf{S}_1^n\right) - P_n^2 \end{aligned}$$

A set B_n resulting as reducing the MSE should penalize large values of $-\mathbf{S}_1^n$ while bearing nearly all the realizations of $\mathbf{S}_1^n$ under the i.i.d. sampling scheme π^{a_n} as n tends to infinity. It should therefore be of the form (nb_n, ∞) for some b_n so that

(a)

$$\lim_{n\to\infty} E_{\pi^{a_n}}\mathbf{1}_{(nb_n,\infty)}\left(\mathbf{S}_1^n\right) = 1$$

and

(b)

$$\lim_{n\to\infty}\sup\frac{E_p\mathbf{1}_{(na_n,\infty)\cap(nb_n,\infty)}\left(\mathbf{S}_1^n\right)\frac{p\left(\mathbf{X}_1^n\right)}{\pi^{a_n}\left(\mathbf{X}_1^n\right)}}{V} < 1$$

which means that the IS sampling density π^{a_n} can lead a MSE defined by

$$MSE(B_n) := E_p\mathbf{1}_{(na_n,\infty)\cap(nb_n,\infty)}\frac{p\left(\mathbf{X}_1^n\right)}{\pi^{a_n}\left(\mathbf{X}_1^n\right)} - P_n^2$$

with a clear gain over the variance indicator. However when $b_n \leq a_n$ (b) does not hold and when $b_n > a_n$ (a) does not hold.

So no reduction of this variance can be obtained by taking into account the properties of the *typical paths* generated under the sampling density: a reduction of the variance is possible only by conditioning on "small" subsets of the sample paths space. On no classes of subsets of $\mathbb{R}^n$ with probability going to 1 under the sampling is it possible to reduce the variability of the estimate, whose rate is definitely proportional to $\sqrt{n}$, imposing a burden of order $L\sqrt{n}\alpha$ in order to achieve a relative efficiency of $\alpha\%$ with respect to P_n.

The MSE of our estimate on a growing class of typical paths

We will evaluate the performance of our estimate under $\mathbf{g}$ since the algorithm envolves technical parameters (typically M); in practice the Monte Carlo approximation introduces no significant bias.

At the contrary to just evidenced hereabove, the procedure which we propose has a small asymptotic variability when evaluated through trimming on classes of subsets of $\mathbb{R}^n$ whose probability goes to 1 under the sampling $\mathbf{g}$. These subsets of $\mathbb{R}^n$ get smaller and smaller as n increases as measured through the MSE of the estimate with respect to the MSE of the classical IS estimate.

We prove the existence of these trimming sets in the present section and state that the resulting gain in terms of the MSE of our estimate is the proper measure of its performance.

These sets are the C_n described in the following Lemma, whose proof is differed to the appendix. For sake of notational simplicity denote ε_n the ε_n' defined in (44).

Lemma 20 *With the just mentioned ε_n, define the family of sets $C_n$in $\mathbb{R}^n$ such that for all x_1^n in C_n,*

$$\left| \frac{\mathfrak{p}_n(x_1^k)}{\mathbf{g}\left(x_1^k\right)} - 1 \right| < \varepsilon_n$$

and

$$\left| \frac{m(t_k)}{a_n} - 1 \right| < \delta_n$$

where t_k is defined through

$$m(t_k) := \frac{n}{n-k} \left(a_n - \frac{s_1^k}{n} \right)$$

and δ_n satisfies

$$\lim_{n\to\infty} \delta_n = 0$$

together with

$$\lim_{n\to\infty} \delta_n a_n \sqrt{n-k} = \infty.$$

Then

$$\lim_{n\to\infty} \mathbf{G}\left(C_n\right) = 1.$$

Furthermore on C_n

$$t_k s(t_k) = a_n \left(1 + o(1)\right). \tag{54}$$

We now prove that our IS algorithm provides a net improvement over the classical IS scheme in terms of Mean Square Error when evaluated on this family of sets.

Define

$$RE\left(\widehat{P_n}\right) = \frac{1}{L} \left(\frac{E_{\mathbf{g}}\left(\mathbf{1}_{C_n}\widehat{P_n}(l)\right)^2}{P_n^2} - 1 \right)$$

$$\widehat{P_n}(l) := \frac{\prod_{i=0}^n p(X_i(l))}{\mathbf{g}(X_1^n(l))} \mathbf{1}_{\mathcal{E}_n}\left(S_1^n(l)\right).$$

We prove that

Proposition 21 *The relative accuracy of the estimate P_n^{IS} is given by*

$$RE(\widehat{P_n}) = \frac{\sqrt{2\pi}\sqrt{n-k-1}}{L} a_n(1+o(1)) \text{ as } n \text{ tends to infinity.}$$

Proof. Denote $E_{\mathfrak{P}_n}$ the expectation with respect to the p.m. $\mathfrak{P}_n$ of $X_1^n(l)$ conditioned upon $\mathcal{E}_n(l) := (S_1^n(l)/n > a_n)$; we omit the index l for brevity. Using the definition of C_n we get

$$\begin{aligned}
E_{\mathbf{g}}\left(\mathbf{1}_{C_n}\widehat{P_n}(l)\right)^2 &= P_n E_{\mathfrak{P}_n} \mathbf{1}_{C_n}(X_1^n) \frac{p(X_1^k)p(X_{k+1}^n)}{\mathbf{g}(X_1^k)\mathbf{g}(X_{k+1}^n/X_1^k)} \\
&\leq P_n(1+\varepsilon_n) E_{\mathfrak{P}_n} \mathbf{1}_{C_n}(X_1^n) \frac{p(X_1^k)}{p(X_1^k/\mathcal{E}_n)} \frac{p(X_{k+1}^n)}{\mathbf{g}(X_{k+1}^n/X_1^k)} \\
&= P_n^2(1+\varepsilon_n) E_{\mathfrak{P}_n} \mathbf{1}_{C_n}(X_1^n) \frac{1}{p(\mathcal{E}_n/X_1^k)} \frac{p(X_{k+1}^n)}{\mathbf{g}(X_{k+1}^n/X_1^k)} \\
&= P_n^2(1+\varepsilon_n)\sqrt{2\pi}\sqrt{n-k-1} E_{\mathfrak{P}_n} \mathbf{1}_{C_n}(X_1^n) t_k s(t_k)(1+o(1)) \\
&= P_n^2 \sqrt{2\pi}\sqrt{n-k-1} a_n(1+o(1)).
\end{aligned}$$

The second line uses $A_{\varepsilon_n}^k$. The third line is Bayes formula. The fourth line is Lemma 3. The fifth line uses (54) and uniformity in Lemma 3, where the conditions in Corollary 6.1.4 of Jensen (1995) are easily checked since, in his notation, $J(\theta) = \mathbb{R}$, condition (i) holds for θ in a neighborhood of 0 (Θ_0 undeed is resticted to such a set in our case), (ii) clearly holds and (iii) is (9). ■

Proposition 22 *When $a_n = n^{-\gamma}$ then under (A) the ratio of the relative efficiencies of the Adaptive IS algorithm with respect to the standard IS scheme is of order $\sqrt{n-k}/\sqrt{n}$.. The same result holds when $a_n = (\log n)^{-\alpha}$*

0.4 Importance Sampling for M-estimators

This Section provides some application of the previous results for some classical types of estimators for which sharp moderate deviation probabilities can be obtained through linear approximations. We follow closely the work by [8]; see also [1].

Let T denote a real valued statistical functional defined on the space M_F, where we assume that T has an Influence Function. Let P be a given p.m. We assume that for all Q in M_F there exists a function g (depending on P) such that

$$\left| T(Q) - T(P) - \int g dQ \right| < \omega\left(N\left(Q-P\right)\right) \tag{55}$$

where N is a seminorm defined on M_F , continuous in the τ_F topology, and ω is a continuous and strictly monotone function which satisfies $\omega(t)/t \to 0$ as $t \to 0$.

The function g is the *Influence Function* of T at P. The class F considered here contains $B(\mathbb{R}) \cup \{g\}$.

Let $\psi(x,t)$ be real valued function defined on $\mathbb{R}^2$ and assume that P satisfies $\int |\psi(x,t)| \, dP < \infty$. Define $T(P)$ as any solution of the equation

$$\int \psi(x,t) dP = 0 \tag{56}$$

if defined. When $P = P_t$ depends upon a real valued parameter t such that

$$T(P_t) = t$$

then T is *Fisher consistent* and the substitution of P_t by P_n in (56), the empirical measure pertaining to an i.i.d. sample with unknown p.m. P_{t_0} provides a consistent estimate of t_0 under appropriate regularity conditions; see [14]. Such estimate is an M-estimator. We assume that all conditions M1 to M5 in [8] hold, which implies that (55) above holds (see [8] Theorem 4.2). Also in this case the function g writes

$$g(x) = \frac{\psi(x,t_0)}{\frac{d}{dt} \int \psi(x,t_0) dP_{t_0}}.$$

The same situation holds for L-estimators,

When (10) holds in M_F it can be checked that a strong MDP holds for $T(P_n)$; Indeed when g belongs to the class F and

$$\lim_{n\to\infty} \left(na_n^2\right)^{-1} \log \left[n P_{t_0} \left(|g(X_1)| > na_n \right) \right] = -\infty$$

then using (55) and (10) it can be proved that the remaining term in $T(P_n) - T(P_{t_0})$ is negligible w.r.t. the linear approximation $\int g(x,t_0) dP_n$ on the moderate deviation scale, as follows from (2.14) and (2.15) in [8]. Furthermore in this case the strong moderate deviation holds for $P_{t_0} \left(|T(P_n) - T\left(P_{t_0}\right)| > a_n \right)$ and

$$\lim_{n\to\infty} \frac{P_{t_0}\left(T(P_n) - T\left(P_{t_0}\right) > a_n\right)}{P_{t_0}\left(\frac{1}{n}\sum_{i=1}^n g(X_i) > a_n\right)} = 1$$

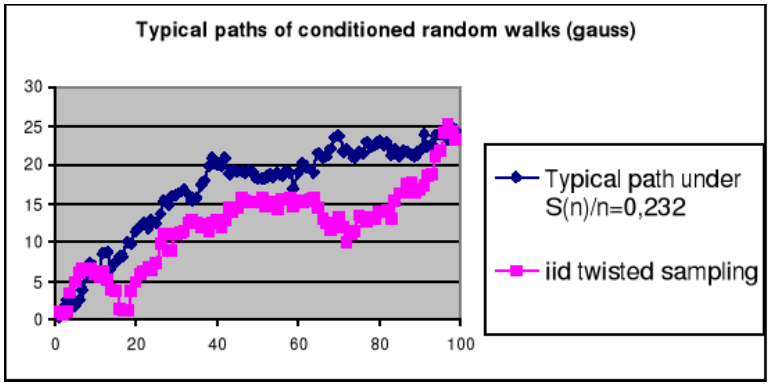
in the range $a_n = n^{-\alpha}, \frac{1}{3} < \alpha < \frac{1}{2}$;. see also Inglot, Kallenberg and Ledwina [11].

0.5 Simulation results

0.5.1 The gaussian case

Typical paths under the final value

This graph illustrates Proposition 8.



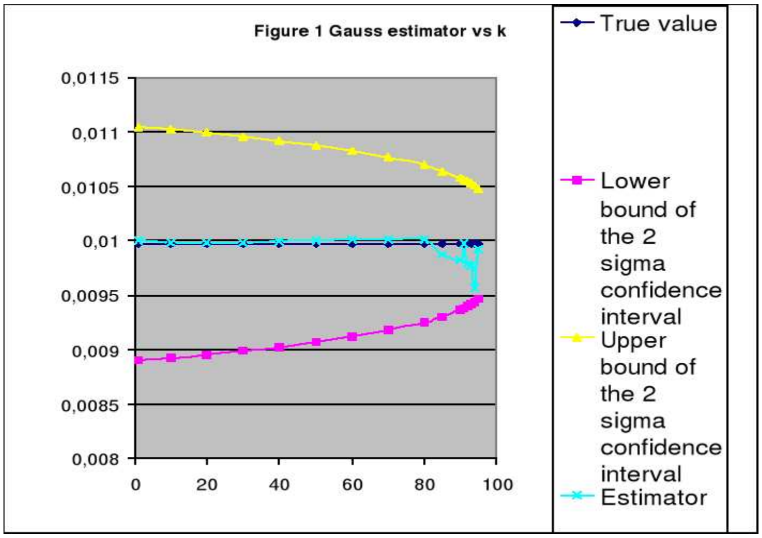


Figure 1Gauss

The graph shows the role of k in the behavior of the estimate. The X_i's are standard normal, $n = 100$ and $P_n = 10^{-2}$. When k is less than 70 the new estimate improves on the classical i.i.d. scheme. A change in M leads no significant change (here $M = 30$). The value of L is $L = 2000$.

Figure 2 Gauss

The graph illustrates the accuracy of the asymptotic results in Propositions 19 and 21. The X_i's are standard normal, $n = 100$,$\mathrm{P}_n = 10^{-2}, k = 60$.

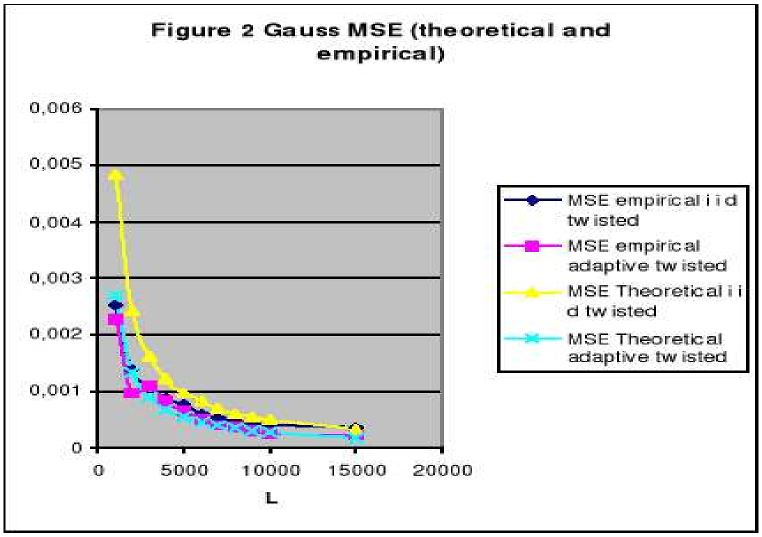


Figure 3 Gauss

The graph is an illustration of Proposition 22. The r.v's X_i's are standard normal, $n = 100$ and $P_n = 10^{-2}$. In ordinate is the ratio of the empirical value of the MSE of the adaptive estimate w.r.t. the empirical MSE of the i.i.d. twisted one. The value of k is $k = 60$; this ratio stabilizes to $\sqrt{n-k}/\sqrt{n}$ for large L, in full accordance with Proposition 22.

0.5.2 The exponential case

typical paths

The graphs above are typical paths under the conditional distribution (with $\mathbf{S}_n/n = 0.239$) and under the i.i.d. sampling with tilted density. The value of n is 100 and the approximation of the conditional density of the random walk is fair up to $k = 80$, as indicated by the fact that the IS estimator of P_n is correct up to $k = 80$, which can be seen as a pertinent indicator.

The random variables $X_i's$ are i.i.d. with exponential distribution with parameter 1 on $(-1, \infty)$. The case treated here is $P\left(\frac{\mathbf{S}_n}{n} > a_n\right) = P_n$ with $n = 100$, $P_n = 0.013887$ and $a_n = 0.232$. These values are computed through a very long run of the standard IS algorithm (with i.i.d. sampling according to the twisted) and are used as a benchmark.The estimates are calculated with $L = 1000$, and $L = 10000$ for $k = 0$, i.e. for the classical i.i.d. twisted sampler (lower values of L lead unstable estimates)

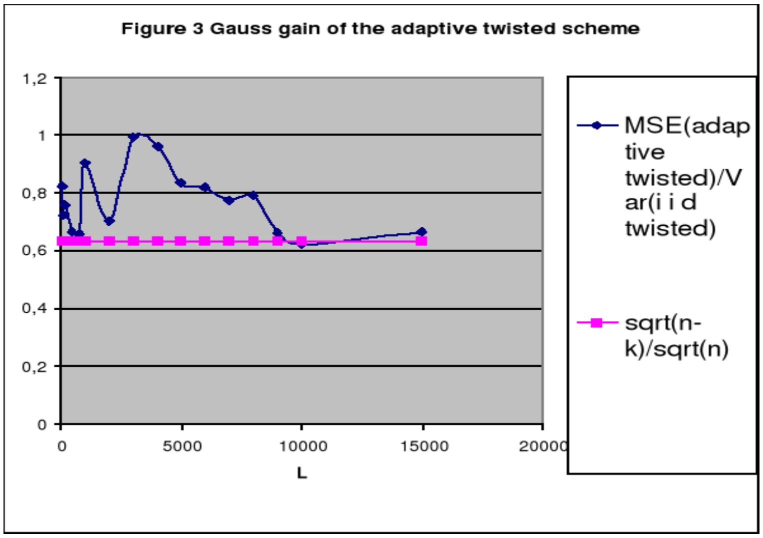
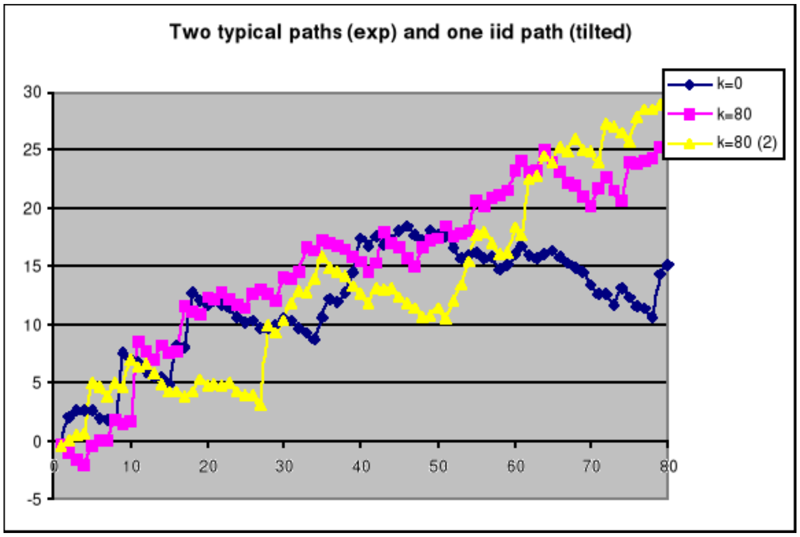
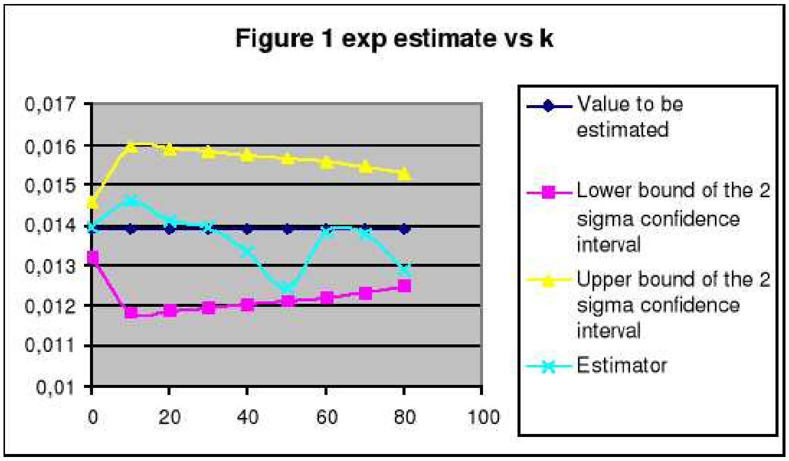


Figure 3 Gauss gain of the adaptive twisted scheme



Two typical paths (exp) and one iid path (tilted)



0.6 Appendix

0.6.1 Proof of Proposition 1

We first state

Lemma 23 *Denote* $q^* := \frac{dQ^*}{d\lambda}$ *(*λ *the Lebesgue measure). Then* $q^*(y) = xyp(y)$

Proof. By Theorem 3.4 (2) in Broniatowski and Keziou (2006) it holds $q^*(y) = (\alpha y + \beta) p(y)$ for some constants α and β. The projection Q^* satisfies both $\int vdQ(v) = x$ and $\int dQ(v) = 0$ which yield $\alpha = x$ and $\beta = 0$. ■

For any set A in $B(\mathbb{R})$, it holds

$$P\left(\mathbf{X}_1 \in A/\left(\mathbf{S}_1^n/n > a_n x\right)\right) = P(A) + a_n x Q^*(A) + o\left(a_n\right). \tag{57}$$

Indeed it holds

$$\begin{aligned}
\frac{1}{na_n^2}\left(P\left(\mathbf{X}_1 \in A/\mathcal{E}_{n,x}\right) - P(A)\right) &= \frac{1}{na_n^2} E\left(1_A(\mathbf{X}) - P(A)/\mathcal{E}_{n,x}\right) \\
&= E\left(\frac{1}{na_n^2}\left(\frac{1}{n}\sum_{i=1}^n 1_A(\mathbf{X}_i) - P(A)\right)/\mathcal{E}_{n,x}\right) \\
&= \int_{-\infty}^0 P\left(\frac{1}{na_n^2}\left(\frac{1}{n}\sum_{i=1}^n 1_A(\mathbf{X}_i) - P(A)\right) < t/\mathcal{E}_{n,x}\right) dt \\
&\quad + \int_0^\infty P\left(\frac{1}{na_n^2}\left(\frac{1}{n}\sum_{i=1}^n 1_A(\mathbf{X}_i) - P(A)\right) > t/\mathcal{E}_{n,x}\right) dt.
\end{aligned}$$

Observe that $\mathcal{E}_{n,x} = \{\mathbf{M}_n \in \Omega_x\}$. Also denote A_t^+ (resp A_t^-) the subset of $M(\mathbb{R})$ defined through $A_t^+ := \left\{Q \in M(\mathbb{R}) : Q(\mathbb{R}) = 0, \int 1_A(v)dQ(v) \geq t\right\}$, resp $A_t^- := \left\{Q \in M(\mathbb{R}) : Q(\mathbb{R}) = 0, \int 1_A(v)dQ(v) < t\right\}$. Using Bayes formula and the above moderate deviation result (10) it follows that for any measurable set G in $M(\mathbb{R})$

$$\lim_{n\to\infty} \Pr\left(M_n \in G/M_n \in \Omega_x\right) = \begin{array}{l} 1 \text{ if } Q^* \text{ belongs to } G \\ 0 \text{ otherwise} \end{array}$$

Proof. For any positive (resp. negative) t it then holds $\lim_{n\to\infty} \Pr\left(\mathbf{M}_n \in A_t^+/\mathbf{M}_n \in \Omega_x\right) = 1$ if $t < Q^*(A)$ and $Q^*(A) > 0$ (resp $\lim_{n\to\infty} \Pr\left(\mathbf{M}_n \in A_t^-/\mathbf{M}_n \in \Omega_x\right) = 1$ if $t > Q^*(A)$ and $Q^*(A) < 0$), which is to say, going to the limit in n, that $\lim_{n\to\infty} \frac{1}{a_n}\left(P\left(\mathbf{X}_1 \in A/\mathcal{E}_{n,x}\right) - P(A)\right) = \int_{-Q^{*-}(A)}^{0} dt + \int_0^{Q^{*+}(A)} dt$ where $Q^* = Q^{*+} - Q^{*-}$ is the Lebesgue decomposition of Q^*. This closes the proof of (57). A second order expansion of $\pi^{a_n x}(y)$ in a neighborhood of $t = 0$ yields

$$\pi^{a_n x}(y) = (1 + a_n x y + a_n^2 x^2 g_n(y))p(y).$$

Hence for all Borel set A it holds $\int_A \pi^{a_n x}(y)dy = P(A) + a_n x Q^*(A) + a_n^2 x^2 \int_A g_n(y))p(y)dy$. Since both $\int_A \pi^{a_n x}(y)dy$ tends to $P(A)$ and Q^* is a finite measure it follows that $a_n^2 x^2 \int_A g_n(y))p(y)dy$ tends to 0. ■

0.6.2 Two Lemmas pertaining to the partial sum under its final value

We now state two lemmas which describe some functions of the random vector $\mathbf{X}_1^n$ conditioned on $\mathcal{E}_n$.

Lemma 24 *Assume that (A) holds. Then for all i between* 1 *and* k

$$\frac{n}{n-i}\left(a_n - \frac{\mathbf{S}_1^i}{n}\right) = a_n + O_{\mathfrak{P}_n}\left(\frac{1}{\sqrt{n-i}}\right).$$

Proof. Select s in (a_n, b_n) and denote P_n^s the p.m on $\mathbb{R}^n$ conditioned on $(\mathbf{S}_1^n = ns)$ It holds

$$\sqrt{n-i}\left(m_{i,n} - a_n\right) = \sqrt{n-i}\left(\frac{\mathbf{S}_{i+1}^n}{n-i} - s\right) + \sqrt{n-i}\left(a_n - s\right).$$

We prove that for $m = n - i$

$$var_{P_n^s}\left(\sqrt{m}\left(\frac{\mathbf{S}_1^m}{m} - s\right)\right) = O(1)$$

as $m \to \infty$ where $var_{P_n^s} Z$ denotes the variance of Z conditionally on $\left(\frac{\mathbf{S}_1^n}{n} = s\right)$. Integrating with respect to the distribution of $\mathbf{S}_1^n$ conditioned upon $\mathcal{E}_n$ concludes the proof. Using ■

$$p_n^s(\mathbf{X}_1=x)=\frac{p_{\mathbf{S}_2^n}\left(ns-x\right)p_{\mathbf{X}_1}(x)}{p_{\mathbf{S}_1^n}\left(ns\right)}=\frac{\pi^t_{\mathbf{S}_2^n}\left(ns-x\right)\pi^t_{\mathbf{X}_1}(x)}{\pi^t_{\mathbf{S}_1^n}\left(ns\right)}$$

with $m(t)=s$, normalizing both $\pi^t_{\mathbf{S}_2^n}\left(ns-x\right)$ and $\pi^t_{\mathbf{S}_1^n}\left(ns\right)$ and making use of a first order Edgeworth expansion in those expressions yields

$$E_{P_n^s}\left(\mathbf{X}_1\right)=s+0\left(\frac{1}{n}\right)$$

and

$$E_{P_n^s}\left(\mathbf{X}_1^2\right)=s^2(t)+s^2+0\left(\frac{1}{n}\right).$$

With a similar development for the joint density $\mathfrak{p}_n(\mathbf{X}_1=x,\mathbf{X}_2=y)$, using the same tilted distribution π^t it readily follows that

$$E_{P_n^s}\left(\mathbf{X}_1\mathbf{X}_2\right)=s^2+0\left(\frac{1}{n}\right).$$

Since

$$var_{P_n^s}\mathbf{S}_1^m=m(m-1)E_{P_n^s}\left(\mathbf{X}_1\mathbf{X}_2\right)+mE_{P_n^s}\left(\mathbf{X}_1^2\right)-m^2E_{P_n^s}\left(\mathbf{X}_1\right)^2$$

it follows that when m/n tends to 0, then $var_{P_n^s}\mathbf{S}_1^m=m\left(1+o(1)\right)$. Since $m\leq n-k$ this amounts to

$$\lim_{n\to\infty}\frac{n-k}{n}=0.$$

Integration with respect to the distribution of $\mathbf{S}_1^n$ conditioned upon $\mathcal{E}_n$ and splitting the integeral on (a_n,a_n+c_n) and (a_n+c_n,∞), using (C2) concludes the proof.

Remark 25 *It can be proved that*

$$\sqrt{m}\left(\frac{\mathbf{S}_1^m}{m}-a_n\right)\Rightarrow N(0,1)\ \textit{when}\ m/n\to 0$$

conditionally on $\left(\mathbf{S}_1^n/n>a_n\right)$. *This result is to be compared with the Gibbs principle for moderate deviations stated in the Introduction which assets that for fixed* m *the joint distribution of* $(\mathbf{X}_1,...,\mathbf{X}_m)$ *conditioned upon* $\mathcal{E}_n$ *converges weakly , as* $n\to\infty$, *to the joint distribution of* m *r.v's* $\mathbf{X}_1^*,...,\mathbf{X}_m^*$ *which are independent copies of* $\mathbf{X}^*$ *.The above result says that even for sequences depending upon* n, *we may replace the original* m *variables by the* m *independent tilted ones when exploring the behavior of* S_1^m *under* $\mathcal{E}_n$, *since* $\sqrt{m}\left(\frac{\mathbf{S}_1^m}{m}-a_n\right)$ *shares the same limit distribution.*

We also need the order of magnitude of $\max\left(\mathbf{X}_1,...,\mathbf{X}_k\right)$ under $\mathfrak{P}_n$ which is stated in the following result.

Lemma 26 *It holds for all k between 1 and n*

$$\max\left(\mathbf{X}_1,...,\mathbf{X}_k\right) = O_{\mathfrak{P}_n}(\log n).$$

Let s such that $na_n \leq s \leq a_n$ $+c_n$. Denote P_n^s the probability measure of $\mathbf{X}_1^n$ given the the value of $\mathbf{S}_1^n = s$. Since

$$\mathfrak{P}_n\left(\max\left(\mathbf{X}_1,...,\mathbf{X}_k\right) > t\right) = \int_{na_n}^{\infty} P_n^s\left(\max\left(\mathbf{X}_1,...,\mathbf{X}_k\right) > t\right) p\left(\mathbf{S}_1^n = s/\mathcal{E}_n\right) ds$$

we first state the order of magnitude of $\max\left(\mathbf{X}_1,...,\mathbf{X}_k\right)$ under P_n^s in the next Lemma.

Lemma 27 *For all k between 1 and $n, \max\left(\mathbf{X}_1,...,\mathbf{X}_k\right) = O_{P_n^s}\left(\log k\right).$*

Proof. Define $\tau := s/n$. For all t it holds

$$\begin{aligned}
P_n^s\left(\max\left(\mathbf{X}_1,...,\mathbf{X}_k\right) > t\right) &\leq kP_n^s\left(\mathbf{X}_n > t\right) \\
&= k\int_t^{\infty} p(\mathbf{X}_n = u/\mathbf{S}_1^n = s)p(\mathbf{S}_1^n = s/\mathcal{E}_n)du \\
&= k\int_t^{\infty} \pi^\tau\left(\mathbf{X}_n = u\right)\frac{\pi^\tau\left(\mathbf{S}_1^{n-1} = s-u\right)}{\pi^\tau\left(\mathbf{S}_1^n = s\right)}du.
\end{aligned}$$

Center and normalize both $\mathbf{S}_1^n$ and $\mathbf{S}_1^{n-1}$with respect to the density π^τ in the last line above, denoting $\overline{\pi_n^\tau}$ the density of $\overline{\mathbf{S}_1^n} := \left(\mathbf{S}_1^n - n\tau\right)/s_\tau\sqrt{n}$ when $\mathbf{X}$ has density π^τ with mean τ and variance s_τ^2, we get

$$\begin{aligned}
P_n^s\left(\max\left(\mathbf{X}_1,...,\mathbf{X}_k\right) > t\right) \leq\ & k\frac{\sqrt{n}}{\sqrt{n-1}}\int_t^{\infty} \pi^\tau\left(\mathbf{X}_n = u\right) \\
& \frac{\overline{\pi_{n-1}^\tau}\left(\overline{\mathbf{S}_1^{n-1}} = (n\tau - u - (n-1)\tau))/\left(s_\tau\sqrt{n-1}\right)\right)}{\overline{\pi_n^\tau}\left(\overline{\mathbf{S}_1^n} = 0\right)}du.
\end{aligned}$$

Under the sequence of densities π^τ the triangular array $\left(\mathbf{X}_1,...,\mathbf{X}_n\right)$ obeys a first order Edgeworth expansion

$$\begin{aligned}
P_n^s\left(\max\left(\mathbf{X}_1,...,\mathbf{X}_k\right) > t\right) &\leq k\frac{\sqrt{n}}{\sqrt{n-1}}\int_t^{\infty} \pi^\tau\left(\mathbf{X}_n = u\right) \\
&\quad \frac{\mathfrak{n}\left((\tau - u)/s_\tau\sqrt{n-1}\right)\mathbf{P}\left(u,i,n\right) + o(1)}{\mathfrak{n}\left(0\right) + o(1)}du \\
&\leq kCst\int_t^{\infty} \pi^\tau\left(\mathbf{X}_n = u\right)du.
\end{aligned}$$

for some constant Cst independent of n and τ and where

$$\mathbf{P}\left(u,i,n\right) := 1 + P_3\left((\tau - u)/s_\tau\sqrt{n-1}\right)$$

where $P_3(x) = \frac{\mu_3^{(\tau)}}{6\left(\sigma^{(\tau)}\right)^3}\left(x^3 - 3x\right)$ is the third Hermite polynomial; $\left(\sigma^{(\tau)}\right)^2$ and $\mu_3^{(\tau)}$ are the second and third centered moments of π^τ. We used uniformity upon u in the remaining term of the Edgeworth expansions. Let t_τ such that $m(t_\tau) = \tau$. Making use of Chernoff Inequality

$$\begin{aligned} P_n^s\left(\max\left(\mathbf{X}_1, ..., \mathbf{X}_k\right) > t\right) &\leq \frac{k}{\phi(t_\tau)}\int_t^\infty \exp-\left(C - t_\tau\right)u\ du \\ &\leq kCst\frac{\phi(t_\tau + \lambda)}{\phi(t_\tau)}e^{-\lambda t} \end{aligned}$$

for any λ such that $\phi(t_\tau + \lambda)$ is finite.

$$t/\log k \to \infty$$

it holds

$$P_n^s\left(\max\left(\mathbf{X}_1, ..., \mathbf{X}_k\right) < t\right) \to 1,$$

which proves the lemma. ■

We now prove Lemma 26

As above write

$$\begin{aligned} \mathfrak{P}_n\left(\max\left(\mathbf{X}_1, ..., \mathbf{X}_k\right) > t\right) &\leq k\mathfrak{P}_n\left(\mathbf{X}_n > t\right) \\ &\leq k\int_{na_n}^\infty\left(\int_t^\infty \pi^\tau\left(\mathbf{X}_n = u\right)\frac{\pi^\tau\left(\mathbf{S}_1^{n-1} = s - u\right)}{\pi^\tau\left(\mathbf{S}_1^n = s\right)}du\right) \\ &\quad p\left(\mathbf{S}_1^n = s/\mathcal{E}_n\right)ds \end{aligned}$$

where τ is defined as in the above Lemma through $\tau := s/n$. Use the same argument as in Lemma 27 to assess that when $t/\log n$ goes to infinity then the.RHS above tends to 0. This closes the proof.

0.6.3 Proof of Lemma 16

It holds

$$\begin{aligned} \mathfrak{p}_n\left(Y_1^k\right) &= \frac{\int_{a_n}^\infty p_{\mathbf{X}_1^k, \mathbf{S}_n}\left(Y_1^k, t\right)dt}{P\left(\mathcal{E}_n\right)} \\ &= \frac{np_{\mathbf{X}_1^k}(Y_1^k)}{(n-k)P\left(\mathcal{E}_n\right)}\int_{a_n}^\infty p_{\mathbf{S}_{k+1}^n/(n-k)}\left(\frac{n}{(n-k)}\left(t - \frac{\Sigma_1^k}{n}\right)\right)dt. \end{aligned}$$

By Lemma 14 it holds under (C1)

$$\frac{\Sigma_1^n}{n} = a_n + R_n$$

where $R_n := O_{\mathfrak{P}_n}\left(\frac{1}{na_n}\right) > 0$. Denote $\mathbf{S} := \frac{\mathbf{S}_{k+1}^n}{n-k}$. Set

$$I = \frac{\int_b^\infty p_{\mathbf{S}}(u)du}{\int_a^\infty p_{\mathbf{S}}(u)du} = \frac{P\left(\mathbf{S} > b\right)}{P\left(\mathbf{S} > a\right)}$$

with $a := \frac{n}{n-k}\left(a_n - \frac{\Sigma_1^k}{n}\right)$ and $b := \frac{n}{n-k}\left(b_n - \frac{\Sigma_1^k}{n}\right)$. It holds

$$\mathfrak{p}_n\left(Y_1^k\right) = (1+I)\int_{a_n}^{b_n} p\left(Y_1^k/\mathbf{T}=\sigma\right)p\left(\mathbf{T}=\sigma\right)d\sigma.$$

Use Lemma **??** to obtain

$$I = \frac{P\left(\mathbf{S} > \alpha_n + \frac{n}{n-k}c_n\right)}{P\left(\mathbf{S} > \alpha_n\right)}$$

where $\alpha_n := a_n + O_{\mathfrak{P}_n}\left(\frac{1}{\sqrt{n-k}}\right) = a_n\left(1 + o_{\mathfrak{P}_n}(1)\right)$. Use Lemma 3 to obtain

$$I = \left(\exp - nc_na_n\right)\left(\exp\frac{n^2c_n^2}{n-k}\right)$$

which tends to 0 under (C).

0.6.4 Proof of Lemma 20

The approximation in (44) holds only on

$$A_{n,\varepsilon_n} := A_{\varepsilon_n}^k \times \mathbb{R}^{n-k}.$$

In the above display,

$$A_{\varepsilon_n}^k := \left\{x_1^k : \left|\frac{\mathfrak{p}_n(x_1^k)}{\mathbf{g}\left(x_1^k\right)} - 1\right| < \varepsilon_n\right\}.$$

By the above definition

$$\lim_{n\to\infty}\mathfrak{P}_n\left(A_{n,\varepsilon_n}\right) = 1 \tag{58}$$

Note also that

$$\begin{aligned}\mathbf{G}\left(A_{n,\varepsilon_n}\right) \quad : \quad &= \int \mathbf{1}_{A_{n,\varepsilon_n}}(x_1^n)\mathbf{g}\left(x_1^n\right)dx_1^n = \int \mathbf{1}_{A_{\varepsilon_n}^k}(x_1^k)\mathbf{g}\left(x_1^k\right)dx_1^n \\ &\geq \frac{1}{1+\varepsilon_n}\int \mathbf{1}_{A_{\varepsilon_n}^k}(x_1^k)\mathfrak{p}_n(x_1^k)dx_1^k \\ &= \frac{1}{1+\varepsilon_n}\left(1+o(1)\right)\end{aligned}$$

which goes to 1 as n tends to ∞, where we have used Proposition 15. In the above displays $\mathbf{g}\left(x_1^k\right)$ is the density of X_1^k when X_1^n is sampled under $\mathbf{g}$. We have just proved that the sequence of sets A_{n,ε_n} contains roughly all the sample paths X_1^n under the importance sampling density $\mathbf{g}$.

We use the fact that t_k defined through

$$m(t_k) = \frac{n}{n-k}\left(a_n - \frac{\Sigma_1^k}{n}\right)$$

is close to a_n under $\mathfrak{P}_n$ uniformly upon σ in (a_n, b_n).

Let δ_n tend to 0 and $\lim_{n\to\infty} a_n\delta_n\sqrt{n-k} = \infty$ and

$$B_n := \left\{ x_1^n : \left| \frac{m(t_k)}{a_n} - 1 \right| < \delta_n \right\}.$$

We prove that on B_n

$$t_k s(t_k) = a_n \left(1 + o(1)\right) \tag{59}$$

holds.

By Lemma 24 and (C3)

$$\lim_{n\to\infty} \mathfrak{P}_n \left(B_n\right) = 1. \tag{60}$$

There exists δ_n' such that for any x_1^n in B_n

$$\left| \frac{t_k}{a_n} - 1 \right| < \delta_n'. \tag{61}$$

Indeed

$$\left| \frac{m(t_k)}{a_n} - 1 \right| = \left| \frac{t_k\left(1+v_k\right)}{a_n} - 1 \right| < \delta_n$$

and $\lim_{n\to\infty} v_k = 0$. Therefore

$$1 - \frac{v_k t_k}{a_n} - \delta_n < \frac{t_k}{a_n} < 1 - \frac{v_k t_k}{a_n} + \delta_n.$$

Since $\frac{m(t_k)}{a_n}$ is bounded so is $\frac{t_k}{a_n}$ and therefore $\frac{v_k t_k}{a_n} \to 0$ as $n\to\infty$ which implies (61).

Further (61) implies that there exists δ_n" such that

$$\left| \frac{t_k s(t_k)}{a_n} - 1 \right| < \delta_n".$$

Indeed

$$\begin{aligned} \left| \frac{t_k s(t_k)}{a_n} - 1 \right| &= \left| \frac{t_k\left(1+u_k\right)}{a_n} - 1 \right| \\ &\leq \delta_n' + \left(1+\delta_n'\right) u_k = \delta_n" \end{aligned}$$

where $\lim_{n\to\infty} u_k = 0$. Therefore (59) holds.

Define

$$C_n := B_n \cap A_{n,\varepsilon_n}$$

Since

$$\int \mathbf{1}_{C_n}(x_1^n)\mathbf{g}\left(x_1^k\right) dx_1^n \geq \frac{1}{1+\varepsilon_n} \int \mathbf{1}_{C_n}\mathfrak{p}_n(x_1^n) dx_1^n$$

and by (58) and (60)

$$\lim_{n\to\infty} \mathfrak{P}_n \left(C_n\right) = 1$$

we obtain

$$\lim_{n\to\infty} \mathbf{G}\left(C_n\right) = 1.$$

which concludes the proof.

Bibliography

[1] Miguel A. Arcones. Moderate deviations for M-estimators. *Test*, 11(2):465–500, 2002.

[2] P. Barbe and M. Broniatowski. Simulation in exponential families. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 9(3):203–223, 1999.

[3] Michel Broniatowski and Amor Keziou. Minimization of ϕ-divergences on sets of signed measures. *Studia Sci. Math. Hungar.*, 43(4):403–442, 2006.

[4] James Antonio Bucklew. *Introduction to rare event simulation.* Springer Series in Statistics. Springer-Verlag, New York, 2004.

[5] A. de Acosta. Moderate deviations and associated Laplace approximations for sums of independent random vectors. *Trans. Amer. Math. Soc.*, 329(1):357–375, 1992.

[6] P. Diaconis and D. A. Freedman. Conditional limit theorems for exponential families and finite versions of de Finetti's theorem. *J. Theoret. Probab.*, 1(4):381–410, 1988.

[7] M. S. Ermakov. Asymptotically efficient statistical inferences for moderate deviation probabilities. *Teor. Veroyatnost. i Primenen.*, 48(4):676–700, 2003.

[8] Mikhail Ermakov. Importance sampling for simulations of moderate deviation probabilities of statistics. *Statist. Decisions*, 25(4):265–284, 2007.

[9] William Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[10] Cheng-Der Fuh and Inchi Hu. Efficient importance sampling for events of moderate deviations with applications. *Biometrika*, 91(2):471–490, 2004.

[11] Tadeusz Inglot, Wilbert C. M. Kallenberg, and Teresa Ledwina. Strong moderate deviation theorems. *Ann. Probab.*, 20(2):987–1003, 1992.

[12] Jens Ledet Jensen. *Saddlepoint approximations*, volume 16 of *Oxford Statistical Science Series*. The Clarendon Press Oxford University Press, New York, 1995. Oxford Science Publications.

[13] Vol′fgang Rihter. Local limit theorems for large deviations. *Dokl. Akad. Nauk SSSR (N.S.)*, 115:53–56, 1957.

[14] Robert J. Serfling. *Approximation theorems of mathematical statistics.* John Wiley & Sons Inc., New York, 1980. Wiley Series in Probability and Mathematical Statistics.

[15] Jan M. Van Campenhout and Thomas M. Cover. Maximum entropy and conditional probability. *IEEE Trans. Inform. Theory*, 27(4):483–489, 1981.

[16] Sandy L. Zabell. Rates of convergence for conditional expectations. *Ann. Probab.*, 8(5):928–941, 1980.